Box-Jenkins seasonal forecasting.

Prothero, D. L.

Award date:
1975

Awarding institution:
University of Bath

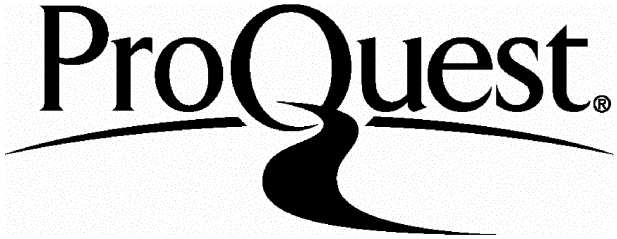
BOX-JENKINS SEASONAL FORECASTING

Submitted by D.L. Prothero
for the degree of Ph.D.
of the University of Bath
1975.

COPYRIGHT

D.L. Prothero

D.L. Prothero



CONTENTS

ABSTRACT

The forecasting procedure recently developed by Professors Box and Jenkins, and described in Box and Jenkins (1970), is based on a class of models (A.R.I.M.A. models) capable of representing a wide range of time series. In this thesis we examine some of the practical problems involved in applying the Box-Jenkins procedure to seasonal time series. A Box-Jenkins analysis of a series of sales figures is described in detail and some of the problems encountered during this analysis are dealt with at length. The topics examined include the application of non-linear transformations in time series analyses and the employment of differencing operators as a means of producing a stationary process. The computation of the unconditional sum of squares when estimating the parameters in an A.R.I.M.A. model and the performance of the Box-Jenkins procedure when applied to series which include deterministic components are also investigated. The A.R.I.M.A. model arising when a time series is considered to be generated by stochastic trend, seasonal and extraneous error components is developed while the interpretation of A.R.I.M.A. models, and their generated forecasts, in terms of the more familiar concepts of trend and seasonality, is explored. A summary of 4 further Box-Jenkins analyses is given, special reference being made to the topics mentioned above. The performance of the Box-Jenkins procedure is compared with that of the method proposed by Winters (1960), on the 5 series included in this thesis.

CHAPTER 1

INTRODUCTION

1.1 Objectives

In business management today, almost every decision made at executive level is based on some kind of forecast. The financial consequences of poor forecasting can therefore be so serious that reliable and detailed forecasts are now regarded as essential in such areas as production planning and stock control. This increasing need for accurate forecasts in business and economics has stimulated the development of a number of new forecasting techniques over the last twenty years or so. The many techniques currently available possess various degrees of complexity, ranging from inspired guesswork to methods based on complicated statistical models.

One particular forecasting technique which has recently aroused a great deal of interest is the method developed by Professors Box and Jenkins and described in Box and Jenkins (1970). Reid (1969) found that this method generally compared very favourably with other univariate forecasting techniques when applied over a large sample of economic time series.

The idea of this thesis originated following an approach by an engineering firm (Company X). This firm supplied a seasonal series, consisting of the monthly sales figures of an engineered product, and forecasts for a lead time of up to 12 months were required. In the light of the promising results obtained by Reid (1969) it was decided to apply the Box-Jenkins procedure to the data of Company X.

The Box-Jenkins forecasting procedure will be described in

Chapter 2 while a detailed account of the analysis of the Company X data will be presented in Chapter 3. The following six chapters will deal with some of the problems encountered during the case study of Chapter 3 and also with certain features of the Box-Jenkins procedure which, it was considered, necessitated further attention. The topics to be covered will include the application of a non-linear transformation to the data prior to performing a Box-Jenkins analysis (Chapter 4), the use of differencing operators as a means of inducing stationarity (Chapter 5) and a deeper look at some of the steps involved in the estimation procedure employed by Box and Jenkins (Chapter 6). In Chapters 7 and 8 it will be assumed that any given series can be decomposed into trend, seasonal and error components. The use of the Box-Jenkins procedure on series which include deterministic trend and seasonal components will be examined in Chapter 7 while in Chapter 8 models involving stochastic trend and seasonal components will be related to the class of models on which the Box-Jenkins procedure is based. The interpretation of this latter class of models will be discussed in Chapter 9.

In order to gain further experience with the Box-Jenkins procedure, it was applied to a further 4 seasonal time series. Chapter 10 will report on these analyses and also on how the Box-Jenkins procedure compared with the method proposed by Winters (1960), on these 4 series and the Company X series.

Finally, in Chapter 11, the material included in this thesis will be summarised, conclusions will be drawn and areas of further research will be suggested.

At various stages it will prove necessary to refer to other forecasting methods. For this reason we shall begin by briefly reviewing some alternative forecasting techniques. Section 1.2 will be devoted to this review while some of the considerations governing the choice of the correct forecasting technique will be discussed in

Section 1.3. An account of two comprehensive empirical comparisons of univariate forecasting techniques will be given in Section 1.4. The first was undertaken by Reid (1969), the second by Newbold and Granger (1974).

1.2 A Review of Forecasting Techniques

In this section it is assumed that the many forecasting procedures can be divided into three main categories: qualitative techniques, univariate techniques and multivariate techniques. A similar classification has been adopted by Chambers et al. (1971) and Chatfield (1974).

1.2.1 Qualitative Techniques

A qualitative forecasting technique is defined to be one which uses qualitative data to produce quantitative forecasts. Qualitative data is a term used to describe data derived from a variety of sources. Expert opinion, human judgement and market research are just three examples of sources of qualitative data. Others are described by Chambers et al. (1971).

Qualitative techniques have proved most useful in situations where no historical data are available. In cases when a new product is introduced into the market, Green and Harrison (1973) have suggested a Bayesian approach.

1.2.2 Univariate Techniques

The techniques outlined in this subsection derive forecasts which are based entirely on current and past values of the variable to be forecasted. More formally, given a time series consisting of observations X_t ($t = 1,2,3,\ldots,n$) made at discrete equally spaced intervals of time, a univariate technique will produce a forecast of some future value, $X_{n+\ell}$, using only the past observations

X_t (t = 1,2,3,...,(n-1)) and the current observation X_n. This forecast will be denoted by $\hat{X}_n(\ell)$ where n refers to the time base (or origin) from which the forecast is made and ℓ to the distance into the future one is forecasting. The latter symbol is termed the lead time.

Any forecasting problem can be considered to fall into one of three categories: short-term, medium-term or long-term forecasting. The exact definition of what constitutes these three ranges depends on the area in which one is working. In relation to sales forecasting the short-term is usually up to about nine months, the medium-term is the next two or three years and the long-term is anything in excess of this. Examples of the meaning of short, medium and long-term in respect to other fields are given by Kendall (1973, page 115).

Many forecasting procedures could be included in this subsection, some simple, others much more sophisticated. Almost all the univariate techniques can be termed fully automatic in the sense that once a computer programme has been written, forecasts can be generated without further human intervention. The technique developed by Professors Box and Jenkins is however a notable exception.

Brief accounts of the more important univariate techniques are now given.

Trend Projections

This technique is most applicable to the problem of long-term forecasting. Essentially the method of trend projections involves fitting a trend curve (e.g. polynomial, exponential) to past data and extrapolating. A full discussion of the use of trend curves for forecasting is given by Gregg et al.(1964) and Harrison and Pearce (1972).

The univariate techniques to be reviewed hereafter are generally useful only for short or medium-term forecasting.

Moving Averages

Moving averages have been dealt with at length by Kendall and Stuart (1966, Vol. III) and also by Brown (1963). As an example, for a locally trend-free, non-seasonal series, X_t, the moving average at time t (and hence the forecast for all lead times made at time t) would be a simple (i.e. equally weighted) average of the most recent N observations. The choice of an appropriate value for N has been discussed by Brown (1963).

In practice moving averages are seldom used for forecasting purposes in their own right. Of more importance is the fact that they provide a starting point in the development of more sophisticated procedures.

Simple Exponential Smoothing

The technique of simple exponential smoothing is a logical extension of the method of moving averages and is appropriate only for trend-free non-seasonal data. Instead of using a simple average of past observations, forecasts are computed on the basis of exponentially (or more correctly, geometrically) weighted moving averages (E.W.M.A.'s). Given a series X_t it is easy to show that the E.W.M.A. at time t, m_t, can be expressed as

$$m_t = AX_t + (1-A)m_{t-1} \qquad 1.2.1$$

or

$$m_t = m_{t-1} + Ae_t \qquad 1.2.2$$

where m_{t-1} is the previous value of the E.W.M.A., $e_t = X_t - m_{t-1}$ is the forecast error appropriate to time t and A is termed the smoothing constant ($0 < A < 1$).

The forecast made at time t for all lead times is simply m_t.

Originally advocated by Holt, simple exponential smoothing has a rather limited practical use due to its inability to account for trend and seasonality.

More detailed accounts of simple exponential smoothing can be found in Winters (1960), Brown (1963), Coutie et al. (1964) and Harrison (1965).

Holt-Winters Procedure

This procedure is described fully by Winters (1960) and Coutie et al. (1964). Essentially the technique of simple exponential smoothing is extended to cover time series which exhibit trend and seasonality. For a series possessing a local linear (or additive) trend and a multiplicative seasonal variation, the Holt-Winters forecasting model is based on the following equations:

$$m_t = A\frac{X_t}{s_{t-L}} + (1-A)(m_{t-1} + r_{t-1}) \qquad 1.2.3$$

$$r_t = B(m_t - m_{t-1}) + (1-B)r_{t-1} \qquad 1.2.4$$

$$s_t = C\frac{X_t}{m_t} + (1-C)s_{t-L} \qquad 1.2.5$$

where $X_t (t = 1,2,3,\ldots,n)$ is the given time series,

m_t represents an estimate of the level of the series at time t,

r_t represents the current estimate of the linear trend factor,

s_t represents the estimated seasonal factor appropriate to time t

and L is the period of the seasonal cycle.

A, B and C are all smoothing constants (or parameters) which can be estimated by computing the sum of squared forecast errors over a grid of values for A, B and C and choosing those values which minimise this quantity.

Initial values for m_t, r_t and s_t are determined from the first H

observations of the series in question. These values are usually taken to be

$$m_1 = \frac{1}{L} \sum_{t=1}^{L} X_t \qquad 1.2.6$$

$$r_1 = \frac{\sum_{t=H-L+1}^{H} X_t - \sum_{t=1}^{L} X_t}{L(H-L)} \qquad 1.2.7$$

and $$s_j = \frac{L}{H} \sum_{i=1}^{H/L} s_{i,j} \quad , \quad j = 1,2,3,\ldots,L \qquad 1.2.8$$

where H is chosen so that H/L is an integer and

$$s_{i,j} = \frac{X_t}{m_1 - \left[\frac{L+1}{2} - j\right] r_1} \quad , \quad \begin{matrix} i = 1,2,3,\ldots H/L \\ j = 1,2,3,\ldots,L \end{matrix} .$$

$s_{i,j}$ is the estimated seasonal factor for the j^{th} period in the i^{th} cycle and so $t = j + (i-1)L$. The seasonals obtained from equation (1.2.8) may have to be normalised to ensure that they sum to L.

The ℓ-step ahead forecast made at time t is given by

$$\hat{X}_t(\ell) = (m_t + \ell r_t)\, s_{t-L+\ell} \quad , \quad \ell = 1,2,3,\ldots,L \qquad 1.2.9$$

The forecasting model defined by equations (1.2.3), (1.2.4), (1.2.5) and (1.2.9) can be modified in an obvious way to account for additive seasonal variations or indeed cases in which no seasonal pattern is present.

Brown's Method (General Exponential Smoothing)

As the alternative name implies, this method is a generalised form of simple exponential smoothing. Brown (1963) assumes that any given series, X_t $(t = 1,2,3,\ldots,n)$, can be described locally by a linear combination of m functions of time, viz.,

$$X_t = \sum_{i=1}^{m} \alpha_i(t)\, f_i(t) + e_t \qquad 1.2.10$$

where e_t is a random error. The components $f_i(t)$ may for example be polynomials, exponentials or sinusoidal functions. The coefficients at time t, $\alpha_i(t)$, are estimated by minimising the sum of discounted squared errors

$$\sum_{j=0}^{t-1} \beta^j e_{t-j}^2 \quad , \quad 0 < \beta < 1 \, .$$

The forecast made at time t, for a future observation ℓ steps ahead is

$$\hat{X}_t(\ell) = \sum_{i=1}^{m} \hat{\alpha}_i(t)\, f_i(\ell) \qquad 1.2.11$$

where $\hat{\alpha}_i(t)(i = 1,2,3,\ldots,m)$ is the estimate for $\alpha_i(t)$.

Brown (1963) goes on to show that under certain conditions, updating formulae for the $\hat{\alpha}_i(t)$'s can be derived.

The main feature of Brown's model is that it involves only one smoothing parameter, β. This can be contrasted with the Holt-Winters model which relies on two parameters for non-seasonal data and three parameters for seasonal data. In fact Harrison (1965) suggests that for seasonal forecasting Brown's method is not satisfactory since a suitable choice of the single parameter β cannot be made.

Harrison's Seasonal Method

The criticism of Brown's seasonal method has already been mentioned. The Holt-Winters technique for dealing with seasonals also has its drawbacks, the chief one being that each seasonal factor is updated only once every complete cycle. An improvement proposed by Harrison (1965) is to smooth the most recent seasonal factors $s_t, s_{t-1}, s_{t-2}, \ldots, s_{t-L+1}$, (obtained using equation (1.2.5)) by a Fourier analysis. The Fourier coefficients are then estimated by

$$c_k = \frac{2}{L} \sum_{j=1}^{L} s_j \cos k \lambda_j$$

$$d_k = \frac{2}{L} \sum_{j=1}^{L} s_j \sin k \lambda_j \qquad 1.2.12$$

for $k = 1,2,3,\ldots, L/2$, where $\lambda_j = \frac{2(j-1)\pi}{L} - \pi$.

The smoothed seasonal factors $\bar{s}_{t-L+j}$ $(j = 1,2,3,\ldots,L)$ are given by

$$\bar{s}_{t-L+j} = 1 + \sum_{\text{sig } k} (c_k \cos k \lambda_j + d_k \sin k \lambda_j) \qquad 1.2.13$$

The symbol $\sum_{\text{sig } k}$ denotes the summation over significant harmonics. The smoothed seasonals can be conveniently updated using formulae developed by Harrison (1965).

Step-wise Autoregression

This technique has been suggested by Kendall (1973) and Newbold and Granger (1974). It is in some ways similar to the Box-Jenkins procedure and for this reason a description will be delayed until the next chapter.

Harrison-Stevens Bayesian Approach

This approach developed by Harrison and Stevens (1971) modifies the Holt-Winters linear growth model to take into account the probabilities π_i $(i = 1,2,3,\ldots,m)$ that the system is in one of m states at any given time. For example, the process may be in a state of "step change" at a certain time, implying that a permanent "jump" in the level of the series occurred at that time.

The main disadvantage of the procedure is that the computing time is generally much greater than for most other univariate techniques.

1.2.3 Multivariate Techniques

A multivariate forecasting technique is one in which forecasts are computed on the basis of the current and past values of the variable being forecasted together with current and past values of other variables which are related to this variable. Much care has to be exercised in selecting the related variables and for this reason multivariate forecasting procedures are generally more expensive and take longer to develop than their univariate counterparts.

Multiple Regression Techniques

Multiple linear regression techniques involve regressing the variable to be forecasted, $X_{1,t}$, on certain lagged values of some explanatory variables $X_{2,t}$, $X_{3,t}$, $X_{4,t}$, ..., $X_{m,t}$ and also possibly on past values of $X_{1,t}$. The problem of determining the right explanatory variables and the lags at which each should enter the regression equation has been discussed by Kendall (1973). An example of forecasting using lagged relationships can be found in the paper by Coen et al. (1969).

The subject of forecasting using regression techniques has proved to be the centre of a good deal of argument. For various opinions on the subject, the reader is referred to Brown (1963, page 77), Coen et al. (1969), Box and Newbold (1971), Granger and Newbold (1972) and Kendall (1973).

Econometric Models

Detailed accounts of econometric models and their use in forecasting are generally best provided by economic texts, e.g. Bridge (1971), Christ (1966). The description given by Kendall (1973, page 141) should prove adequate for any reference made to econometric models in this thesis.

Box-Jenkins Input-Output Model

The Box-Jenkins univariate procedure (see Chapter 2) is basically extended to cover the situation in which an "output" variable Y_t is related to some "input" variable X_t. In general several input variables can be considered. A full account of this approach is given by Box and Jenkins (1968, 1970).

1.3 Selection of Forecasting Techniques

Chambers et al. (1971) discuss the problem of choosing the forecasting technique most appropriate to any given situation. Many factors need to be considered when making this choice, the most important of which are summarised below:

(a) The amount of money a company or individual is prepared to spend on a forecasting technique.

(b) The time available for making the forecasts.

(c) The context in which the forecast is to be used.

(d) The availability of historical data.

(e) The accuracy required from the forecast.

(f) The distance into the future for which the forecast is required.

(g) The number of items to be forecasted.

(h) Whether the data are seasonal or non-seasonal.

An examination of how the selection of a forecasting technique may be restricted on account of the factors (a) to (h) is now carried out.

In the absence of historical data the forecasting technique must be chosen from the qualitative class. Assuming this is so, if little money is available and forecasts are required quickly, then the relevant technique will, of necessity, be based on little more than guesswork. However, given time and sufficient money, more sophisticated techniques such as market research can be used. In such a situation, cruder techniques could still be employed but the relatively

low cost and short time required by the latter methods must be balanced against the more accurate forecasts (particularly in the short-term) which the more involved techniques would be expected to yield.

If historical data are available then qualitative methods would rarely be used alone. The choice of technique would generally rest between univariate and multivariate procedures. We shall first assume that the time and money needed to set up a multivariate model are not available. If long term forecasts are required then the method of trend projections is the only really appropriate technique. For short or medium-term forecasting, the field is much wider. A comprehensive coverage of the factors governing the choice of a univariate technique for short or medium-term forecasting will not be given in this section. Instead reference should be made to the conclusions arrived at by Reid (1969) and Newbold and Granger (1974). More details of the comparative studies undertaken by these authors will be given in Section 1.4.

Multivariate forecasting techniques become candidates for selection when there are few financial restrictions and when forecasts are not required with any great haste. If used properly multivariate techniques should generally produce forecasts which are at least as accurate as those derived from univariate procedures. However under certain conditions (see Kendall (1973, page 151)) it may be unwise to employ a multivariate technique, while the context in which the computed forecast is to be used should also be considered. A forecast required as a standard or "norm" or a forecast used as a target value could be generated quite effectively from a univariate procedure. Indeed it is in such a context that univariate procedures can prove most useful. On the other hand if the forecast is to be used for planning purposes or decision making then a multivariate procedure may well be called for.

So far the applicability of individual forecasting techniques has been discussed. Where two or more techniques are appropriate Bates and Granger (1969) suggest combining forecasts. This idea would appear to be useful when for example both univariate and multivariate forecasts are available or when two univariate forecasts have been computed (see Section 1.4). Dickinson (1973) has however shown that certain problems do exist in this area.

We have seen how the selection of a forecasting technique can be restricted by the factors mentioned earlier. Nevertheless, in most cases, there will be a number of applicable techniques. Given such an occurrence, the choice will often rest between a relatively cheap robust technique on the one hand and a more costly sophisticated technique on the other hand. Only experience will decide whether the extra expense involved in the latter can be justified.

1.4 The Studies of Reid and Newbold and Granger

When confronted with the problem of forecasting the sales of the engineering firm, Company X, the question of which technique to employ obviously arose. On the reasoning that experience should be gained with univariate techniques before embarking on the more complicated multivariate techniques, it was decided to employ a univariate procedure. In choosing the appropriate univariate procedure, reference was made to an empirical comparative study undertaken by Reid (1969). A condensed version of this study appears in Reid (1971).

Reid (1969) applied the univariate forecasting techniques proposed by Winters (1960) (referred to as the Holt-Winters procedure), Brown (1963), Harrison (1965) and Box and Jenkins (1970) to about 100 economic time series. Not every technique was applied to every series, e.g. seasonal techniques were only applied to ~~seasonal~~

seasonal series. Most of the data analysed were national or industry-wide aggregates which tend to be less volatile than many individual firms' sales data. Both seasonal (quarterly and monthly) and non-seasonal series, composed of at least 50 observations, were examined.

Each series was divided into two parts. The first part was used to fit the appropriate model which was then employed to generate forecasts over the second part of the series. The forecasts were compared on the basis of a quadratic cost of error function, i.e. by examining the mean squared forecast errors.

Reid found that for all series (seasonal and non-seasonal) the Box-Jenkins procedure generally did better than Holt-Winters which in turn did better than Brown's method. A closer examination of the results for monthly seasonal series revealed that Box-Jenkins again came out on top, followed by Harrison's method (only appropriate for seasonal data) and then Holt-Winters. There was however generally no great difference between the latter two except that Harrison's method performed better on series which had both a very strong seasonal factor and fairly large random fluctuations. Brown's method behaved particularly poorly on seasonal series.

From his experience with the comparative study, Reid constructed a decision tree for choosing the forecasting technique most appropriate to a given set of conditions. The factors governing this choice were stated to be the length of the series available, whether the data are seasonal or not, the importance of essentially unpredictable random components, non-stationarities and the lead time being predicted.

The evidence of Reid's study suggested that it would be worthwhile to apply the Box-Jenkins procedure to the Company X data. At the same time it should be remembered that most of Reid's data were national or industry-wide as opposed to the sales of an individual

firm such as Company X.

Recently, the results of a second comprehensive empirical comparison of univariate forecasting techniques have been published. Newbold and Granger (1974) compared the forecasting performance of Box-Jenkins, Holt-Winters and the method of step-wise autoregression (see Section 2.7) on 106 time series. The collection of series included seasonal and non-seasonal, macro-economic and micro sales data. As in the case of Reid's study, each series was divided into fitting and forecasting periods and the comparison was again based on a quadratic cost of error function.

Some of the conclusions arrived at by Newbold and Granger confirmed the results of Reid (1969). The Box-Jenkins procedure generally outperformed both Holt-Winters and step-wise autoregression. This superiority was most marked for short lead times but rather less so for higher lead times. Overall there was little to choose between Holt-Winters and step-wise autoregression although the former performed somewhat better for higher lead times.

In addition to comparing individual forecasting techniques, Newbold and Granger also considered the combination of pairs of forecasts using the approach of Bates and Granger (1969). The most interesting finding was that over a sample of 80 monthly series, the individual Box-Jenkins forecasts were only slightly better than the combined Holt-Winters and step-wise autoregression forecasts (fully automatic forecasts).

The empirical comparisons discussed in this section are the two most extensive to appear in the literature. Instances in which the performance of the Box-Jenkins procedure has been compared with that of alternative forecasting techniques on a small number of series include Box and Jenkins (1970), Naylor et al. (1972) and Bloomfield (1973).

CHAPTER 2

THE BOX-JENKINS APPROACH TO FORECASTING A SINGLE TIME SERIES

2.1 Introduction

In Chapter 1 an outline of the most commonly used univariate forecasting techniques was given. Many of these techniques were based on the principal of exponential smoothing and were generally appropriate for a particular type of process, e.g. a series possessing a linear trend. More recently, Box and Jenkins (1968, 1970) have proposed a class of models capable of representing a wide variety of time series. The fitting of one of these models to a given set of data and the consequent adaptation to forecasting is generally referred to as the Box-Jenkins forecasting procedure.

This chapter is devoted to an account of the Box-Jenkins forecasting procedure. A fuller description of this procedure appears in Box and Jenkins (1970) while other less detailed accounts can be found in Box and Jenkins (1968), Thompson and Tiao (1971), Naylor et al. (1972) and Newbold (1973a).

All the time series mentioned in this chapter are considered to be composed of observations made at discrete equally spaced intervals of time. Sections 2.2 and 2.3 will deal with the class of models capable of describing, respectively, stationary and non-stationary series. The steps involved in the model fitting process will be explained in Section 2.4 while Section 2.5 will demonstrate how the fitted models are used to generate forecasts. In Section 2.6 the models will be extended to cover seasonal series. Section 2.7 will briefly describe the concept of step-wise autoregression and its similarities to the Box-Jenkins procedure.

2.2 Stationary Time Series Models

A series z_t is said to be stationary to the second order if it possesses a constant mean μ, a constant variance σ^2 and constant autocovariances $\gamma(k)$ ($k = \pm1, \pm2, \pm3,\ldots$). If the z_t's are normally distributed then second order (weak) stationarity is equivalent to strict stationarity as defined for example by Anderson (1971).

Let w_t, w_{t-1}, w_{t-2},... be values of a stationary time series at discrete equally spaced times t, t-1, t-2,... . Suppose also that a_t, a_{t-1}, a_{t-2},..., are uncorrelated random variables, all normally distributed with mean zero and variance σ_a^2. Now the observations w_t, w_{t-1}, w_{t-2},... may be highly correlated. A model is therefore required to transform the series w_t into a series of uncorrelated random variables a_t. The two basic models considered by Box and Jenkins (1970) are the autoregressive model and the finite moving average model. These are both examples of stochastic models (as opposed to deterministic models). The meaning of the term stochastic is explained by Box and Jenkins (1970, page 7).

2.2.1 Autoregressive Model

The model

$$\dot{w}_t - \phi_1 \dot{w}_{t-1} - \phi_2 \dot{w}_{t-2} - \ldots - \phi_p \dot{w}_{t-p} = a_t \qquad 2.2.1$$

where $\dot{w}_t = w_t - \mu$ (μ being the mean of the series w_t), is called an autoregressive model of order p. Box and Jenkins (1970) utilise the operator B defined by

$$B^r \dot{w}_t = \dot{w}_{t-r}$$

to write equation (2.2.1) as

$$\phi_p(B) \dot{w}_t = a_t \qquad 2.2.2$$

where $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$

2.2.2 Moving Average Model

The model

$$\dot{w}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \qquad 2.2.3$$

is called a moving average model of order q. Making use of the B operator leads to

$$\dot{w}_t = \theta_q(B) a_t \qquad 2.2.4$$

where $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$

2.2.3 Mixed Models

A combination of equations (2.2.2) and (2.2.4) gives the model

$$\phi_p(B)\dot{w}_t = \theta_q(B) a_t \qquad 2.2.5$$

which is termed the general mixed autoregressive-moving average model of order (p, q) (A.R.M.A. (p, q)). Substituting $w_t - \mu$ for $\dot{w}_t$ in equation (2.2.5) gives

$$\phi_p(B) w_t = \theta_o + \theta_q(B) a_t \qquad 2.2.6$$

where $\theta_o = \phi_p(1)\mu = (1 - \phi_1 - \phi_2 - \ldots - \phi_p)\mu$.

It is shown in Box Jenkins (1970, page 74) that for equation (2.2.5) to represent a stationary process, the roots of the equation $\phi_p(B) = 0$ (B considered to be the variable) must lie outside the unit circle. The roots of the equation $\theta_q(B) = 0$ are also required to lie outside the unit circle. When this condition is satisfied the model is said to be invertible. The reasons for imposing this latter restriction on the model (2.2.5) have been discussed by Kendall (1971) and Chatfield and Prothero (1973b) in

addition to Box and Jenkins (1970).

2.3 Non-stationary Time Series Models

Many time series encountered in practice are best described by a non-stationary model. For example, sales data often exhibit an upward trend, the mean level of the series changing with time. The Box-Jenkins approach to this particular problem is to difference the non-stationary series until a stationary series results. A stationary model of the type described in Section 2.2 is then fitted to the differenced series.

Suppose that z_t is the original series and that a stationary series w_t is produced after differencing d times. Then we may write

$$w_t = \nabla^d z_t = (1 - B)^d z_t \qquad 2.3.1$$

Substituting $w_t = (1 - B)^d z_t$ in equation (2.2.6) leads to

$$\phi_p(B)(1 - B)^d z_t = \theta_o + \theta_q(B)a_t \qquad 2.3.2$$

which is termed an autoregressive integrated moving average model of order (p, d, q) (A.R.I.M.A. (p, d, q)). The word "integrated" arises from the fact that the non-stationary process z_t is obtained by integrating or summing the stationary process w_t.

In practice p, d and q have often been found to take values 0, 1 or 2.

2.4 Model Fitting

Having introduced a class of models suitable for describing the behaviour of a time series, our attention is now focussed on the problem of fitting these models to a suitably transformed series z_t (t = 1, 2, 3, ..., N) which should include at least 50 observations. Box and Jenkins (1970, Chapters 6, 7 and 8) use

an iterative procedure of identification, estimation and diagnostic checking.

2.4.1 Identification

This stage of the model fitting procedure is concerned with the choice of values for p, d and q in equation (2.3.2) and the calculation of preliminary estimates for the resulting model parameters.

The main tool employed in the identification of an A.R.I.M.A. model is the sample autocorrelation function.

The sample autocorrelation coefficient, $r_w(k)$, at lag k for a series w_t composed of n observations is defined by

$$r_w(k) = \frac{c_w(k)}{c_w(o)} \qquad 2.4.1$$

where $$c_w(k) = \frac{1}{n} \sum_{t=1}^{n-k} (w_t - \bar{w})(w_{t+k} - \bar{w})$$

and $$\bar{w} = \frac{1}{n} \sum_{t=1}^{n} w_t$$

Box and Jenkins (1970, page 33) suggest that the maximum value k should be allowed to take in about $n/4$.

The quantity $r_w(k)$ regarded as a function of k is the sample autocorrelation function.

If the sample autocorrelation function of the series z_t dies out quickly then it is reasonable to assume that z_t is already stationary and no differencing is required, i.e. d = 0. On the other hand, if the sample autocorrelation function is slow to die out non-stationarity exists and z_t is differenced successively until a series $w_t = \nabla^d z_t$ is obtained for which the sample autocorrelation function does die out fairly rapidly. The degree

of differencing required to produce this effect is the value for d in the A.R.I.M.A. model (2.3.2). A deeper examination of the selection of the appropriate differencing operator will be given in Chapter 5.

The initial values for p and q are arrived at by considering the theoretical autocorrelation function, $\rho_w(k)$, of various stationary processes.

The theoretical autocorrelation function for an autoregressive process of order p satisfies the difference equation (see Box and Jenkins (1970, page 54)).

$$\rho_w(k) - \phi_1 \rho_w(k-1) - \phi_2 \rho_w(k-2) - \ldots - \phi_p \rho_w(k-p) = 0 \qquad 2.4.2$$

for $k > 0$.

Equation (2.4.2) has general solution

$$\rho_w(k) = A_1 G_1^k + A_2 G_2^k + \ldots + A_p G_p^k \qquad 2.4.3$$

$G_1, G_2, \ldots, G_p$ being roots of the characteristic equation

$$G^p - \phi_1 G^{p-1} - \phi_2 G^{p-2} - \ldots - \phi_p = 0 \qquad 2.4.4$$

and $A_1, A_2, \ldots, A_p$ are constants.

Thus, in general, the theoretical autocorrelation function of an autoregressive process consists of a mixture of damped exponentials and damped sine waves. The former arise from the real roots of equation (2.4.4) while the latter occur when pairs of roots G_i, G_j are complex. When $p = 1$

$$\rho_w(k) = \phi_1^k$$

and so the autocorrelation function decays geometrically (or exponentially).

For a moving average process of order q Box and Jenkins (1970, page 68) show that

$$\rho_w(k) = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2} & \text{for } k = 1, 2, 3, \ldots, q \\ 0 & \text{for } k > q \end{cases} \qquad 2.4.5$$

When q = 1, all the theoretical autocorrelation coefficients are zero except for $\rho_w(1)$ which from equation (2.4.5) is

$$\rho_w(1) = \frac{-\theta_1}{1 + \theta_1^2} \quad .$$

For an autoregressive-moving average process of order (p, q), Box and Jenkins (1970, page 75) show that the theoretical autocorrelation function satisfies the equation

$$\rho_w(k) - \phi_1\rho_w(k - 1) - \phi_2\rho_w(k - 2) - \ldots - \phi_p\rho_w(k - p) = 0 \qquad 2.4.6$$

for $k \geq q + 1$.

The first q autocorrelation $\rho_w(1), \rho_w(2), \rho_w(3), \ldots, \rho_w(q)$ depend on the values of the q moving average parameters as well as the p autoregressive parameters. If p > q then the whole autocorrelation function will consist of a mixture of damped exponentials and/or waves. However, for $q \geq p$ the first q - p + 1 autocorrelation coefficients will not follow the general pattern.

When p = 1, q = 1, $\rho_w(1)$ will depend on θ_1 and ϕ_1 while for lags greater than or equal to 2 the autocorrelation function is

given by

$$\rho_w(k) = \phi_1 \rho_w(k - 1)$$

or

$$\rho_w(k) = \phi_1^{k-1} \rho_w(1) \qquad 2.4.7$$

i.e. the autocorrelation function decays exponentially after the first lag.

In practice, tentative values for p and q in the A.R.M.A. model of order (p, q) are obtained by inspecting the sample autocorrelation function of the stationary series w_t. This function is then compared with the theoretical autocorrelation function of certain autoregressive-moving average processes. For example, if the sample autocorrelation function exhibited an approximate geometric decay then, using equation (2.4.3), the first order autoregressive model would be identified, i.e. p = 1, q = 0. On the other hand if the sample autocorrelation coefficient at lag 1 was the only coefficient which differed significantly from zero then (from equation (2.4.5)) a first order moving average model would be appropriate.

In addition to the sample autocorrelation function, other tools are available for identifying A.R.I.M.A. models. The partial autocorrelation function has been employed by Box and Jenkins (1970) while Cleveland (1972) has suggested the use of inverse autocorrelations. However, in the latter case satisfactory means of estimating the quantities involved have yet to be discovered.

The identification stage is completed by computing preliminary estimates of the parameters included in the tentative model. These estimates are arrived at by expressing the parameters in terms of the theoretical autocorrelation coefficients and replacing the latter by their sample estimates. For example, it has already been shown that the autocorrelation function of the first order auto-

regressive process is given by

$$\rho_w(k) = \phi_1^k$$

and hence $\rho_w(1) = \phi_1$. Thus the sample autocorrelation coefficient at lag 1, $r_w(1)$, is an initial estimate for the parameter ϕ_1.

2.4.2 Estimation

The next step following the identification procedure is to find efficient estimates for the parameters in the tentative model. The problem of fitting the A.R.I.M.A. model (2.3.2) to the non-stationary zeries z_t is equivalent to that of fitting, to the w_t's, the stationary model (2.2.5).

Under the assumption that the a_t's are normally distributed, the maximum likelihood estimates of the ϕ_i's (i = 1, 2, 3,...,p) and θ_j's (j = 1, 2, 3,...,q) will usually, to a good approximation, be given by minimising the sum of squares

$$S(\phi_1,\phi_2,\phi_3,\ldots,\phi_p,\theta_1,\theta_2,\theta_3,\ldots,\theta_q)$$

$$= \sum_t a_t^2(\phi_1,\phi_2,\phi_3,\ldots,\phi_p,\theta_1,\theta_2,\theta_3,\ldots,\theta_q)$$

i.e. $$S(\underline{\phi},\underline{\theta}) = \sum_t a_t^2(\underline{\phi},\underline{\theta}) \qquad 2.4.8$$

where $\underline{\phi} = (\phi_1,\phi_2,\phi_3,\ldots,\phi_p)$ and $\underline{\theta} = (\theta_1,\theta_2,\theta_3,\ldots,\theta_q)$.

The a_t's are determined for given values of $\underline{\phi}$ and $\underline{\theta}$ by rewriting equation (2.2.5) in the form

$$a_t = \dot{w}_t - \phi_1\dot{w}_{t-1} - \phi_2\dot{w}_{t-2} - \ldots - \phi_p\dot{w}_{t-p}$$

$$+ \theta_1 a_{t-1} + \theta_2 a_{t-2} + \ldots + \theta_q a_{t-q} \qquad 2.4.9$$

Certain initial values of the w_t's and the a_t's must however be specified and there are a number of ways in which this can be done. If the series is quite long and no roots of $\phi_p(B) = 0$ are close to the boundary of the unit circle then initial unknown values can be set equal to their expectations, zero in the case of the a_t's and μ for the w_t's. Given that w_t has n observations, the sum of squares defined by equation (2.4.8) will become

$$S(\underline{\phi},\underline{\theta}) = \sum_{t=1}^{n} a_t^2(\underline{\phi},\underline{\theta}|w_t = \mu,\ a_t = 0 \text{ for } t \leq 0) \qquad 2.4.10$$

An improvement on the above approach is to set $a_t = 0$ for $t \leq p$ and then calculate the a_t's for $t = p + 1$ onwards using equation (2.4.9). The sum of squares is then

$$S(\underline{\phi},\underline{\theta}) = \sum_{t=p+1}^{n} a_t^2(\underline{\phi},\underline{\theta}|a_t = 0 \text{ for } t \leq p) \qquad 2.4.11$$

The loss of information in summing over n - p values of the a_t's instead of n values will be unimportant for long series. However, for short series or seasonal data this method is not satisfactory.

Box and Jenkins (1970, page 211) recommend the use of back-forecasting to calculate values for the unknown a_t's and w_t's. This process is now described.

If the w_t's are generated by the model (2.2.5) then they are also generated by the model

$$\phi_p(F)\dot{w}_t = \theta_q(F)e_t \qquad 2.4.12$$

(see Box and Jenkins (1970, page 199)) where e_t, e_{t-1}, e_{t-2},... is

a sequence of independent random variables, each normally distributed with zero mean and variance $\sigma^2_{e.} = \sigma^2_a$. The operator F defined by

$$F^r w_t = w_{t+r}$$

is termed the forward shift operator.

Model (2.4.12) may be expressed as

$$e_t = \dot{w}_t - \phi_1 \dot{w}_{t+1} - \phi_2 \dot{w}_{t+2} - \ldots - \phi_p \dot{w}_{t+p}$$

$$+ \theta_1 e_{t+1} + \theta_2 e_{t+2} + \ldots + \theta_q e_{t+q} \qquad 2.4.13$$

Letting $[e_t]$ denote the expectation of e_t, conditional on $\underline{\phi}$, $\underline{\theta}$ and $w_1, w_2, w_3, \ldots, w_n$, then algebraically we may write

$$[e_t] = \underset{t}{E}\{e_t | \underline{\phi}, \underline{\theta}, w_1, w_2, w_3, \ldots, w_n\}$$

Taking conditional expectations throughout equation (2.4.13) we get

$$[e_t] = [\dot{w}_t] - \phi_1[\dot{w}_{t+1}] - \phi_2[\dot{w}_{t+2}] - \ldots - \phi_p[\dot{w}_{t+p}]$$

$$+ \theta_1[e_{t+1}] + \theta_2[e_{t+2}] + \ldots + \theta_q[e_{t+q}] \qquad 2.4.14$$

where

$$[\dot{w}_t] = \dot{w}_t \quad , \quad \text{for } t = 1, 2, 3, \ldots, n$$

$$[e_j] = 0 \quad , \quad \text{for } j \leq 0$$

$$[e_j] = 0 \quad , \quad \text{for } j > n - p$$

Starting with $t = n - p$, the $[e_t]$'s can be evaluated working backwards using equation (2.4.14) recursively. When $[e_1]$ has been determined, the back-forecasts $[\dot{w}_{-j}]$ $(j = 0, 1, 2, \ldots)$ are computed (using equation (2.4.14)) until some point $(j = K)$ is reached after which $[\dot{w}_{-j}] \simeq 0$ i.e. for $j = K, K + 1, K + 2, \ldots$.

The $[a_t]$'s can now be computed for $t = -(K - 1)$ up to $t = n$ by taking conditional expectations in equation (2.4.9).

i.e. $$[a_t] = [\dot{w}_t] - \phi_1[\dot{w}_{t-1}] - \phi_2[\dot{w}_{t-2}] - \ldots - \phi_p[\dot{w}_{t-p}]$$

$$+ \theta_1[a_{t-1}] + \theta_2[a_{t-2}] + \ldots + \theta_q[a_{t-q}] \qquad 2.4.15$$

with

$$[a_{-j}] = 0 \quad , \quad \text{for } j \geq K$$

and hence

$$S(\underline{\phi},\underline{\theta}) = \sum_{t=-(K-1)}^{n} [a_t(\underline{\phi},\underline{\theta})]^2 \qquad 2.4.16$$

The above is called the unconditional sum of squares function while the sum of squares arrived at by the two simpler methods are termed conditional sum of squares functions.

In computing the unconditional sum of squares function it is possible to repeat the cycle involving equations (2.4.14), (2.4.15) and (2.4.16). Box and Jenkins (1970, page 217) say that in practice "a second iterative cycle would almost never be needed." An examination of situations in which more iterative cycles are required will be made in Chapter 6.

If the identified model contains no more than three parameters then the latter can be estimated quite conveniently using a graphical technique. The unconditional sum of squares is calculated over a grid of values for each parameter and plotted against the parameter values. This will lead to curves (for one parameter models) or contours (for two or three parameter models) from which the values

of the parameters producing the minimum sum of squares can be obtained.

Another method by which the parameters can be estimated is the iterative procedure of non-linear least squares estimation. This process is described fully in Box and Jenkins (1970, pages 231-242). However, the same authors emphasise that this technique should only be used when one is satisfied that no anomalies exist in the estimation situation. This point can only be tested by plotting the sum of squares function for each new estimation problem.

2.4.3 Diagnostic Checking

If the least squares estimates of the parameters $\underline{\phi}$ and $\underline{\theta}$ in the model (2.2.5) are denoted by $\hat{\underline{\phi}}$ and $\hat{\underline{\theta}}$ then

$$\hat{\phi}_p(B)\dot{w}_t = \hat{\theta}_q(B)\hat{a}_t \qquad 2.4.17$$

or

$$\hat{a}_t = \hat{\theta}_q^{-1}(B)\ \hat{\phi}_p(B)\dot{w}_t \qquad 2.4.18$$

The sequence of random variables $\hat{a}_t$, $\hat{a}_{t-1}$, $\hat{a}_{t-2}$,..., defined by equation (2.4.18) are known as the residuals.

Most of the checks on the adequacy of the fitted model are applied to the residuals.

If the correct form of model had been assumed and the true values of the parameters $\underline{\phi}$ and $\underline{\theta}$ were known then

$$a_t = \theta_q^{-1}(B)\ \phi_p(B)\dot{w}_t \qquad 2.4.19$$

and the estimated autocorrelations $r_a(k)$ would be uncorrelated and approximately normally distributed with zero mean and variance $\frac{1}{n}$ (Bartlett (1946)). However, for the residuals $\hat{a}_t$, Box and Pierce

(1970) show that the value of $\frac{1}{n^{\frac{1}{2}}}$ for the standard error of the $r_{\hat{a}}(k)$'s is unreliable at low lags. It tends to underestimate the significance of departures from zero correlations.

Rather than use the above criteria for individual values of the estimated autocorrelation function $r_{\hat{a}}(k)$, Box and Jenkins (1970) suggest a test on the first K estimated autocorrelations considered as a whole. A typical value for K, for non-seasonal data, is 20. Box and Pierce (1970) show that the statistic

$$Q = n \sum_{k=1}^{K} r_{\hat{a}}^2(k)$$

is approximately distributed as χ^2 (K - p - q) for an A.R.M.A. process of order (p, q). Reference to the appropriate percentage points of a χ^2 distribution gives some clue to the adequacy of the proposed model.

While tests on the sample autocorrelation function of the residuals will to a certain extent detect non-randomness, they may not be very sensitive to periodicities. The presence of the latter should be made apparent by examining the periodogram which is defined by

$$I(f_i) = \frac{2}{n}\left[\left(\sum_{t=1}^{n} a_t \cos 2\pi f_i t\right)^2 + \left(\sum_{t=1}^{n} a_t \sin 2\pi f_i t\right)^2\right] \quad 2.4.20$$

where a_t (t = 1, 2, 3,...,n) is the time series under consideration and f_i $(= \frac{i}{n})$ is the frequency. Note that the frequencies f_i are assumed to be harmonics of the fundamental frequency. If the frequencies were allowed to vary continuously in the range 0 - 0.5 cycles then I(f) would be referred to as the sample spectrum (see for example Jenkins and Watts (1968)).

The function $C(f_j)$ defined by

$$C(f_j) = \frac{\sum_{i=1}^{j} I(f_i)}{n\, s^2}$$

where s^2 is an estimate of σ_a^2, is called the normalised cumulative periodogram. If the a_t's are uncorrelated random variables, normally distributed with zero mean and variance σ_a^2 then the normalised cumulative periodogram for the series, plotted against f_j, would consist of points scattered about a straight line. In the presence of, for example, seasonality "humps" will appear at various frequencies. In practice the normalised cumulative periodogram will of course be computed for the residuals $\hat{a}_t$.

Limit lines can be included on both sides of the theoretical line in order to detect possible inadequacies. Again this is discussed by Box and Jenkins (1970, page 297).

Apart from diagnostic checks applied to the residuals, other tests of a model's suitability can be performed. These include examining the need for a further parameter (overfitting) and looking into the possibility that the parameter values change over a period of time.

Should any of the diagnostic checks detect some inadequacy in the original model then another model is identified and the iterative procedure of identification, estimation and diagnostic checking is repeated until a suitable model is found.

2.5 Forecasting

In this section we describe how the A.R.I.M.A. model introduced in Section 2.3 can be adapted for forecasting purposes.

If we assume (without loss of generality) that the stationary series $w_t = \nabla^d z_t$ has a zero mean then employing the non-stationary operator $\phi^*_{p+d}(B)$, defined by Box and Jenkins (1970, page 88) as

$$\phi^*_{p+d}(B) = 1 - \phi^*_1 B - \phi^*_2 B - \ldots - \phi^*_{p+d} B^{p+d} = \phi_p(B)(1 - B)^d$$

the A.R.I.M.A. model (2.3.2.) becomes

$$\phi^*_{p+d}(B) z_t = \theta_q(B) a_t \qquad 2.5.1$$

Thus for some future value $z_{t+\ell}$ we can write

$$z_{t+\ell} = \phi^*_1 z_{t+\ell-1} + \phi^*_2 z_{t+\ell-2} + \ldots + \phi^*_{p+d} z_{t+\ell-p-d}$$

$$+ a_{t+\ell} - \theta_1 a_{t+\ell-1} - \theta_2 a_{t+\ell-2} - \ldots - \theta_q a_{t+\ell-q} \qquad 2.5.2$$

If $\hat{z}_t(\ell)$ denotes the optimal forecast for lead time ℓ from origin t, Box and Jenkins (1970, page 127) show that

$$\hat{z}_t(\ell) = \underset{t}{E}[z_{t+\ell}] \qquad 2.5.3$$

where $\underset{t}{E}[z_{t+\ell}]$ denotes the conditional expectation of $z_{t+\ell}$ given knowledge of all the z's up to time t.

Further,

$$a_{t+1} = z_{t+1} - \hat{z}_t(1) \qquad 2.5.4$$

so that the required forecast $\hat{z}_t(\ell)$ can be obtained by taking conditional expectations throughout equation (2.5.2), making use of the fact that

$$\left.\begin{aligned} \underset{t}{E}[z_{t+j}] &= \hat{z}_t(j) \\ \underset{t}{E}[a_{t+j}] &= 0 \end{aligned}\right\} \quad \text{for } j = 1, 2, 3, \ldots$$

and

$$\left.\begin{aligned} \underset{t}{E}[z_{t-j}] &= z_{t-j} \\ \underset{t}{E}[a_{t-j}] &= z_{t-j} - \hat{z}_{t-j-1}(1) \end{aligned}\right\} \text{ for } j = 0, 1, 2, \ldots$$

In practice the true values of $\phi_1^*, \phi_2^*, \phi_3^*, \ldots, \phi_{p+d}^*$, $\theta_1, \theta_2, \theta_3, \ldots, \theta_q$ and the a_t's would not be known and they would be replaced by the values estimated by the methods described in Section 2.4.

It is also useful to express the A.R.I.M.A. model (2.5.1) in the form of an infinite moving average viz.

$$\begin{aligned} z_t &= (\psi_0 + \psi_1 B + \psi_2 B^2 + \ldots)a_t \\ &= \psi(B)a_t \end{aligned} \qquad 2.5.5$$

where the ψ weights satisfy

$$\phi_p(B)\,(1 - B)^d\,\psi(B) = \theta_q(B) \qquad 2.5.6$$

Employing this representation we have

$$\begin{aligned} \hat{z}_t(\ell) &= \underset{t}{E}[z_{t+\ell}] = \psi_0 \underset{t}{E}[a_{t+\ell}] + \psi_1 \underset{t}{E}[a_{t+\ell-1}] \\ &\quad + \psi_2 \underset{t}{E}[a_{t+\ell-2}] + \ldots \\ &= \psi_\ell\, a_t + \psi_{\ell+1}\, a_{t-1} + \psi_{\ell+2}\, a_{t-2} + \ldots \end{aligned} \qquad 2.5.7$$

and so

$$z_{t+\ell} = \psi_0 a_{t+\ell} + \psi_1 a_{t+\ell-1} + \psi_2 a_{t+\ell-2} + \cdots$$

$$= \hat{z}_t(\ell) + \sum_{j=0}^{\ell-1} \psi_j a_{t+\ell-j} \qquad 2.5.8$$

The distribution of $z_{t+\ell}$, conditional on z_1, z_2, z_3,...,z_t, $p(z_{t+\ell}|z_t, z_{t-1},\ldots,z_1)$, will therefore have a mean $\hat{z}_t(\ell)$ and variance $(1 + \sum_{j=1}^{\ell-1} \psi_j^2)\sigma_a^2$ ($\psi_o = 1$). In addition, providing that the a_t's are normally distributed with zero mean and variance σ_a^2 then the distribution $p(z_{t+\ell}|z_t, z_{t-1},\ldots,z_1)$ will also be normal. Hence approximate $(1 - \alpha) \times 100\%$ probability limits for $\hat{z}_t(\ell)$ are

$$\hat{z}_t(\ell) \pm u_{\alpha/2} \{1 + \sum_{j=1}^{\ell-1} \psi_j^2\}^{\frac{1}{2}} s_a$$

where $u_{\alpha/2}$ is the appropriate percentage point of the unit normal distribution and s_a^2 is the sample variance of the a_t's.

Finally, it is of interest to note that by making use of equation (2.5.7), Box and Jenkins (1970, page 134) show that a forecast made at time t for lead time ℓ can be updated when the observation z_{t+1} becomes available, using the formula

$$\hat{z}_{t+1}(\ell-1) = \hat{z}_t(\ell) + \psi_{\ell-1} a_{t+1} \qquad 2.5.9$$

where $a_{t+1} = z_{t+1} - \hat{z}_t(1)$

2.6 Seasonal Time Series

In the preceeding sections we have considered the application of the Box-Jenkins procedure to non-seasonal series. Box and Jenkins (1970, Chapter 9) also propose a class of models for describing seasonal data. To illustrate how seasonality is taken into account we will assume that z_t is a seasonal series of monthly

sales figures.

An A.R.I.M.A. model of order (P, D, Q) is fitted to successive sales for one particular month, e.g. September. This model takes the form

$$\Phi_P(B^{12})\ \nabla_{12}^{D}\ z_t = \Theta_Q(B^{12})\ \alpha_t \qquad 2.6.1$$

where $\Phi_P(B^{12}) = 1 - \Phi_1 B^{12} - \Phi_2 B^{24} - \ldots - \Phi_P B^{12P}$

$\Theta_Q(B^{12}) = 1 - \Theta_1 B^{12} - \Theta_2 B^{24} - \ldots - \Theta_Q B^{12Q}$

and $\nabla_{12}\ z_t = z_t - z_{t-12}$

D is the degree of differencing required to reduce the series $z_t, z_{t-12}, z_{t-24}, \ldots$ to stationarity.

Model (2.6.1) thus relates the current September sales to previous September sales and current and past errors $\alpha_t, \alpha_{t-12}, \alpha_{t-24}, \ldots$. For August the equivalent model would be

$$\Phi_P(B^{12})\ \nabla_{12}^{D}\ z_{t-1} = \Theta_Q(B^{12})\ \alpha_{t-1} \qquad 2.6.2.$$

and similarly for the other months. Box and Jenkins (1970, page 304) consider that it is reasonable to assume that the parameters $\underline{\Phi} = (\Phi_1, \Phi_2, \Phi_3, \ldots, \Phi_P)$ and $\underline{\Theta} = (\Theta_1, \Theta_2, \Theta_3, \ldots, \Theta_Q)$ are the same for all months.

The error term α_t from equation (2.6.1) could not be expected to be uncorrelated with the error term α_{t-1} from equation (2.6.2). A further A.R.I.M.A. model of order (p, d, q) is therefore fitted to the series $\alpha_t, \alpha_{t-1}, \alpha_{t-2}, \ldots$. Hence

$$\phi_p(B)\ \nabla^d\ \alpha_t = \theta_q(B)\ a_t \qquad 2.6.3$$

where $\phi_p(B)$, $\theta_q(B)$ and d have been defined in Sections 2.2 and 2.3.

Combining equations (2.6.1) and (2.6.3) leads to the model

$$\phi_p(B)\ \Phi_P(B^{12})\ \nabla^d\ \nabla_{12}^D z_t = \theta_q(B)\ \Theta_Q(B^{12})\ a_t \qquad 2.6.4$$

For a series possessing a seasonal cycle of period s equation (2.6.4) can be modified to

$$\phi_p(B)\ \Phi_P(B^s)\ \nabla^d \nabla_s^D z_t = \theta_q(B)\ \Theta_Q(B^s)\ a_t \qquad 2.6.5$$

which is termed a general multiplicative seasonal model of order $(p, d, q) \times (P, D, Q)_s$.

The model fitting process is essentially the same as that for non-seasonal models.

2.7 The Method of Step-wise Autoregression

This method of forecasting suggested by Newbold and Granger (1974) is based on the autoregressive models introduced in Section (2.2.1). In most cases the first differences $w_t = \nabla z_t$ of the given series z_t are analysed. A model of the form

$$w_t = \mu + \sum_{j=1}^{M} \phi_j\ w_{t-j} + a_t \qquad 2.7.1$$

where a_t is a white poise process, is then assumed. Typical values for M are 10 for non-seasonal data and 25 for monthly data.

At the first step of the model building procedure the lagged value w_{t-j} which contributes most to explaining the variation in w_t is introduced. The next lagged value to be included is the one which most improves the fit obtained after the first step. This

process is continued until no significant improvement in fit occurs at the introduction of further lagged values. At this stage lagged values introduced earlier, which no longer contribute significantly to the fit, are dropped.

Forecasts are obtained by projecting forward equation (2.7.1).

We now make a brief comparision of the method of step-wise autoregression with the Box-Jenkins procedure.

Firstly, the model (2.7.1), as it stands, is the A.R.I.M.A. model of order (M, 1, 0). It is thus a special case of the general A.R.I.M.A. model of order (p, d, q). The absence of moving average terms necessitates a large value for M and so model (2.7.1) does not generally provide a parsimonious representation of the series w_t. Although a number of parameters will be eliminated during the model fitting process, the final model will still generally include more parameters than the A.R.I.M.A. model identified using the Box-Jenkins procedure. However, the inconvenience of working with a model which includes a comparatively large number of parameters is offset to some extent by the fact that it is much easier to estimate autoregressive parameters than moving average parameters.

Treated purely from a routine forecasting point of view, the important difference between step-wise autoregression and the Box-Jenkins procedure is that the former approach is fully automatic while the latter is not. Unlike the Box-Jenkins technique, step-wise autoregression does not involve an inspection of the sample autocorrelation function in order to identify the appropriate model and once the value for M has been specified, forecasts can be generated without further human intervention.

CHAPTER 3

AN EXAMPLE OF A BOX-JENKINS ANALYSIS

3.1 Introduction

The Box-Jenkins procedure, described in Chapter 2, was applied to a seasonal series of monthly sales figures. This chapter will be devoted to a step-by-step account of this analysis. The data will be tabulated and discussed in Section 3.2 while Sections 3.3 to 3.6 will deal respectively with the identification, estimation, diagnostic checking and forecasting stages of the Box-Jenkins procedure. Some alternative models will be examined in Section 3.7. In Section 3.8 general remarks will be made concerning the performance of the Box-Jenkins procedure on this particular set of data.

Most of the computations were carried out using a set of Box-Jenkins forecasting programmes included in the I.C.L. computer package at the University of Bath. However, at certain stages, the approach followed by the I.C.L. programmes differed from that set out in Box and Jenkins (1970). This difficulty was overcome by writing a number of additional individual programmes.

Much of the material contained in this chapter is included in a published paper by Chatfield and Prothero (1973 a).

3.2 The Data

As mentioned in Section 1.1, the data were supplied by an engineering firm (Company X) who required sales forecasts for an engineered product for a lead time of up to 12 months. Monthly observations were available from January 1965 to May 1971. The data are tabulated below and plotted in Figure 3.1.

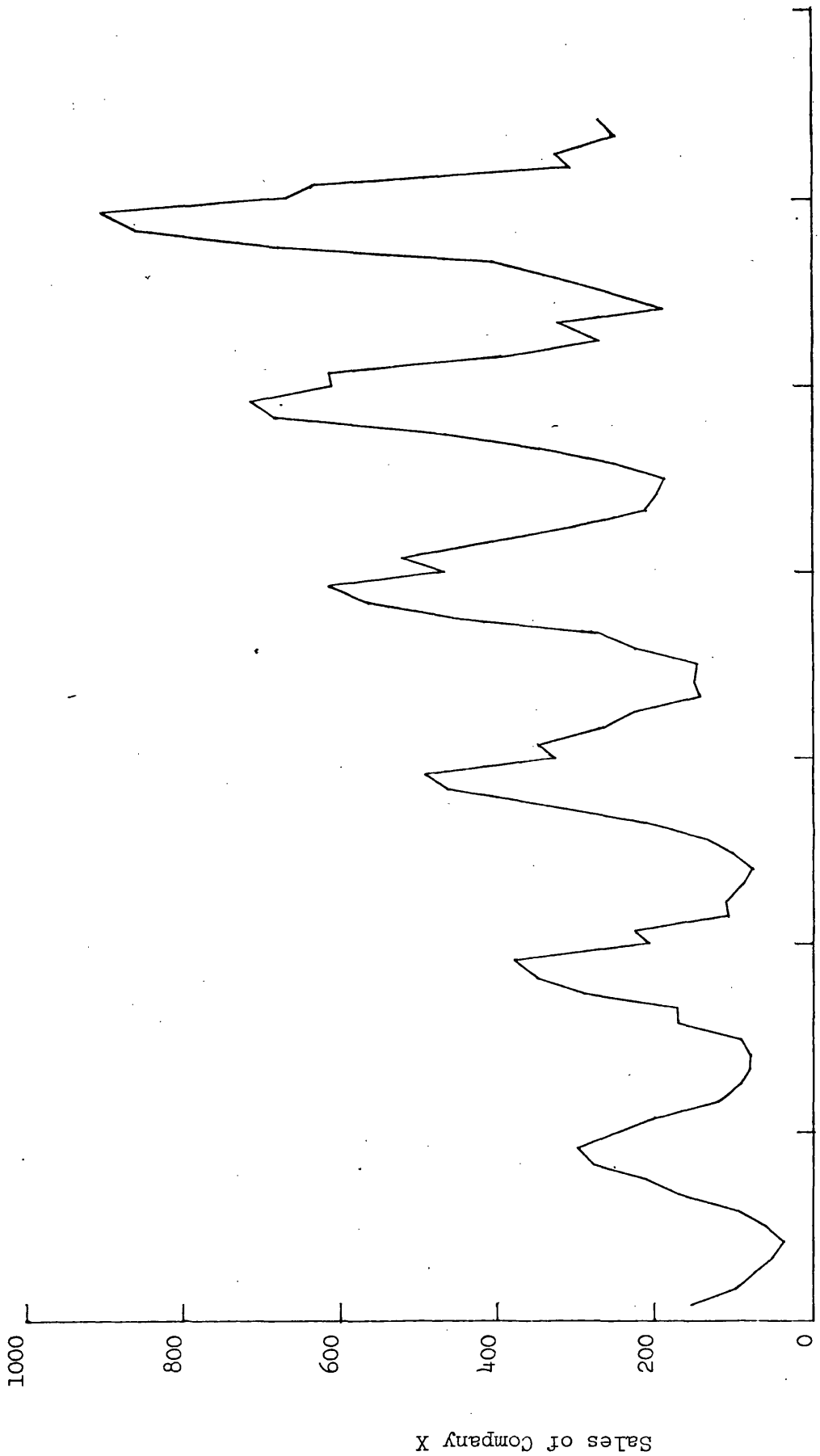


FIGURE 3.1

Sales of Company X, January 1965 - May 1971

Table 3.1 Sales of Company X, January 1965 - May 1971

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
1965	154	96	73	49	36	59	95	169	210	278	298	245
1966	200	118	90	79	78	91	167	169	289	347	375	203
1967	223	104	107	85	75	99	135	211	335	460	488	326
1968	346	261	224	141	148	145	223	272	445	560	612	467
1969	518	404	300	210	196	186	247	343	464	680	711	610
1970	613	392	273	322	189	257	324	404	677	858	895	664
1971	628	308	324	248	272							

The series can be seen to possess a definite upward trend and a marked seasonal pattern. The amplitude of the seasonal cycle is roughly proportional to the level of the series, indicating a multiplicative seasonal effect. At this stage the range of transformations proposed by Box and Cox (1964) was not considered and a logarithmic transformation was applied to the data. The transformed data are shown in Figure 3.2. The seasonal effect is now approximately constant although the trough in the first years data is rather on the low side.

In order to examine the seasonal pattern more closely, the logarithms of the sales for each month were plotted individually (Figure 3.3). These trend lines turn out to be roughly linear and parallel, indicating that a logarithmic transformation is reasonable.

More discussion on the choice of transformations will appear in Chapter 4.

3.3 Identification

Let the observed sales at time t be denoted by X_t and the transformed value by z_t, where

$$z_t = \log_{10} X_t \qquad 3.3.1$$

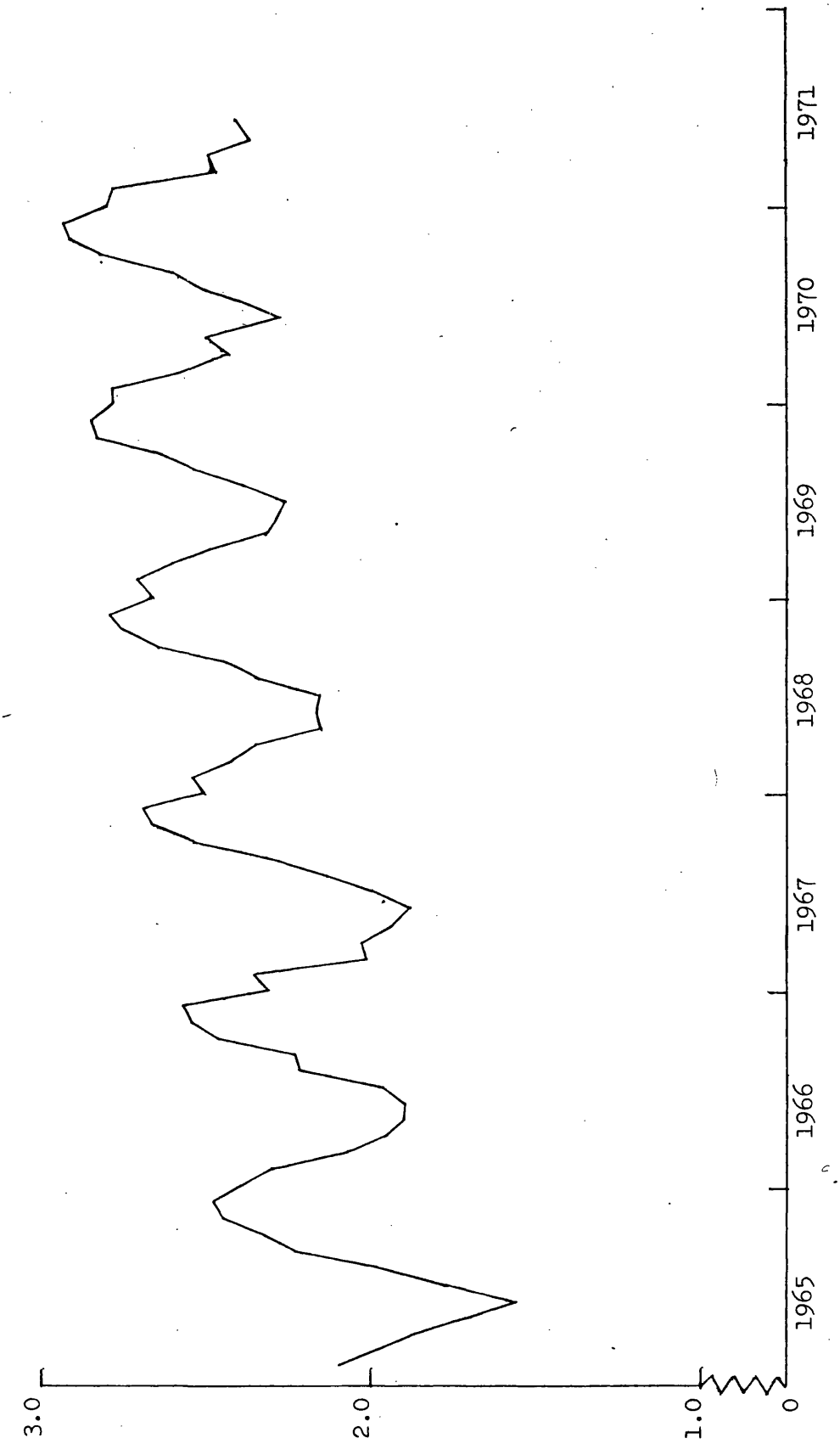


FIGURE 3.2

Logarithms of Sales of Company X, January 1965 - May 1971

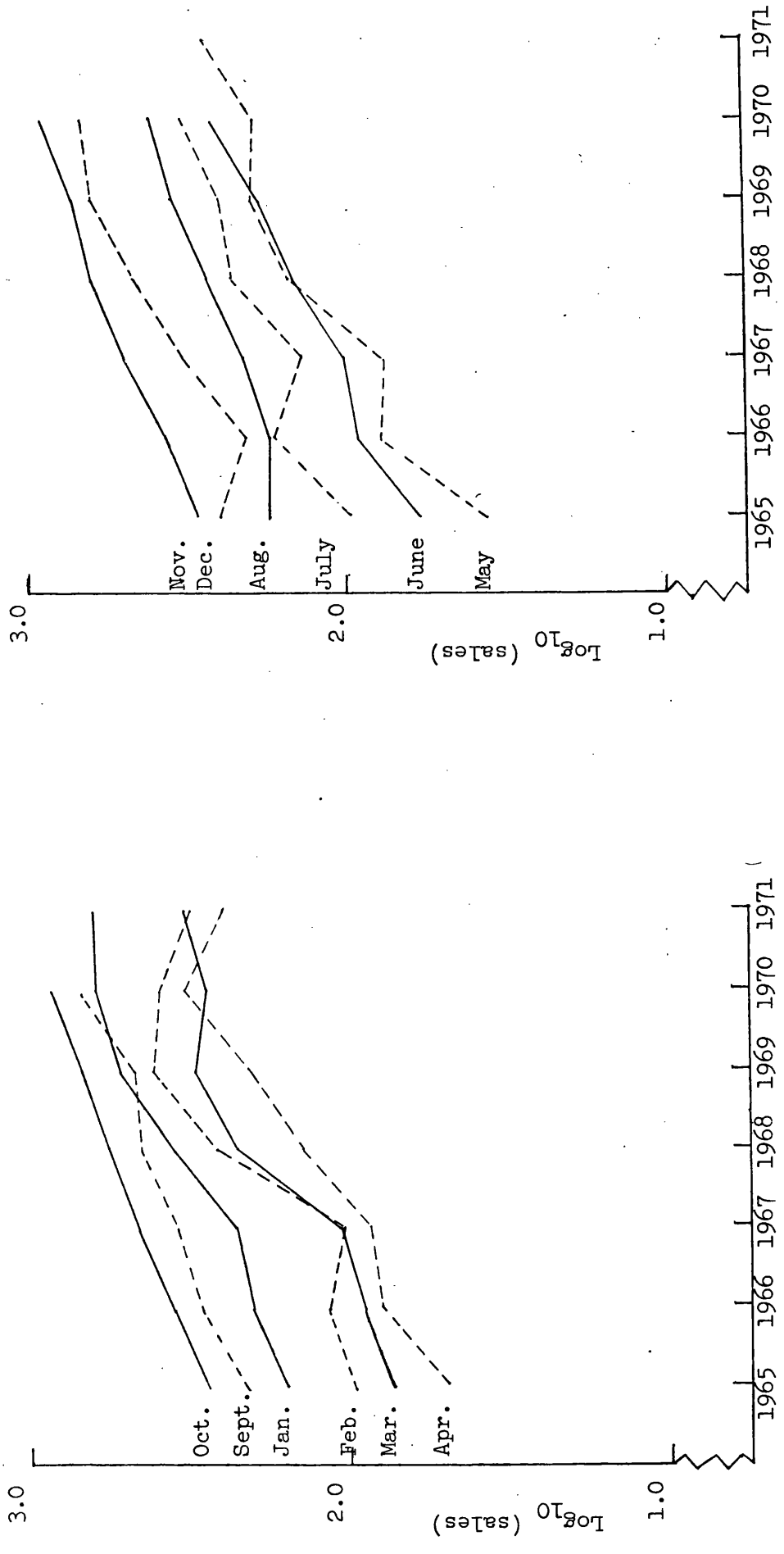


FIGURE 3.3

Log_{10}(sales) for each individual month

We seek to select a model from the class of general multiplicative seasonal models, defined in Section 2.6, which will adequately represent the series z_t. Thus, suitable values for p, d, q, P, D and Q in the model

$$\phi_p(B)\ \Phi_P(B^{12})\ \nabla^d \nabla_{12}^D z_t = \theta_q(B)\ \Theta_Q(B^{12}) a_t \qquad 3.3.2$$

must be chosen. The notation employed in this chapter is the same as that defined in Chapter 2.

3.3.1 Differencing to Attain Stationarity

The first stage of the identification procedure is to determine the degree of differencing necessary to transform the non-stationary series z_t into a stationary series w_t. Following the approach described in Section 2.4.1 the sample autocorrelation functions for various differences of the series z_t were examined. The sample autocorrelation functions for z_t, ∇z_t, $\nabla_{12} z_t$ and $\nabla \nabla_{12} z_t$ are given in Table 3.2 and plotted in Figures 3.4(a), (b), (c) and (d).

The autocorrelation functions for both z_t and ∇z_t show a strong cycle with period 12, the peaks occuring at lags 12 and 24 and the troughs at lags 6 and 18. This suggests that both z_t and ∇z_t are non-stationary. The series $\nabla_{12} z_t$ possesses an autocorrelation function much more like that of a stationary series but a sequence of positive correlations is followed by a long sequence of negative correlations (with one exception) implying that some degree of non-stationarity still exists. On the other hand, the autocorrelations for $\nabla \nabla_{12} z_t$ appear to be quite consistent with those for a stationary series. If the series $\nabla \nabla_{12} z_t$ was random then the standard error of each autocorrelation coefficient would be approximately $\frac{1}{\sqrt{64}} = 0.125$ (see Bartlett (1964)). Thus the only "significant" autocorrelations are those at lags 1, 2, 10, 11 and 12.

Table 3.2 Sample autocorrelation function of $\nabla^d \nabla^D z_t$ for d = 0,1, and D = 0,1

Series	Lags	Autocorrelations											
z_t	1-12	0.90	0.72	0.45	0.19	-0.01	-0.09	-0.06	0.08	0.27	0.47	0.62	0.67
	13-24	0.61	0.46	0.25	0.04	-0.13	-0.20	-0.19	-0.08	0.07	0.24	0.36	0.42
∇z_t	1-12	0.35	0.40	-0.07	-0.30	-0.57	-0.50	-0.54	-0.26	0.02	0.24	0.46	0.59
	13-24	0.37	0.26	0.00	-0.22	-0.44	-0.44	-0.42	-0.21	-0.05	0.24	0.27	0.49
$\nabla_{12} z_t$	1-12	0.45	0.56	0.26	0.21	0.10	0.03	-0.15	-0.14	-0.11	-0.19	0.02	-0.26
	13-24	-0.13	-0.23	-0.15	-0.12	-0.08	-0.10	-0.03	-0.01	0.02	0.01	-0.06	-0.11
$\nabla\nabla_{12} z_t$	1-12	-0.58	0.36	-0.22	0.05	-0.05	0.10	-0.17	-0.02	0.10	-0.26	0.44	-0.36
	13-24	0.19	-0.14	0.04	-0.01	0.05	-0.08	0.04	0.00	0.04	0.05	-0.02	-0.07

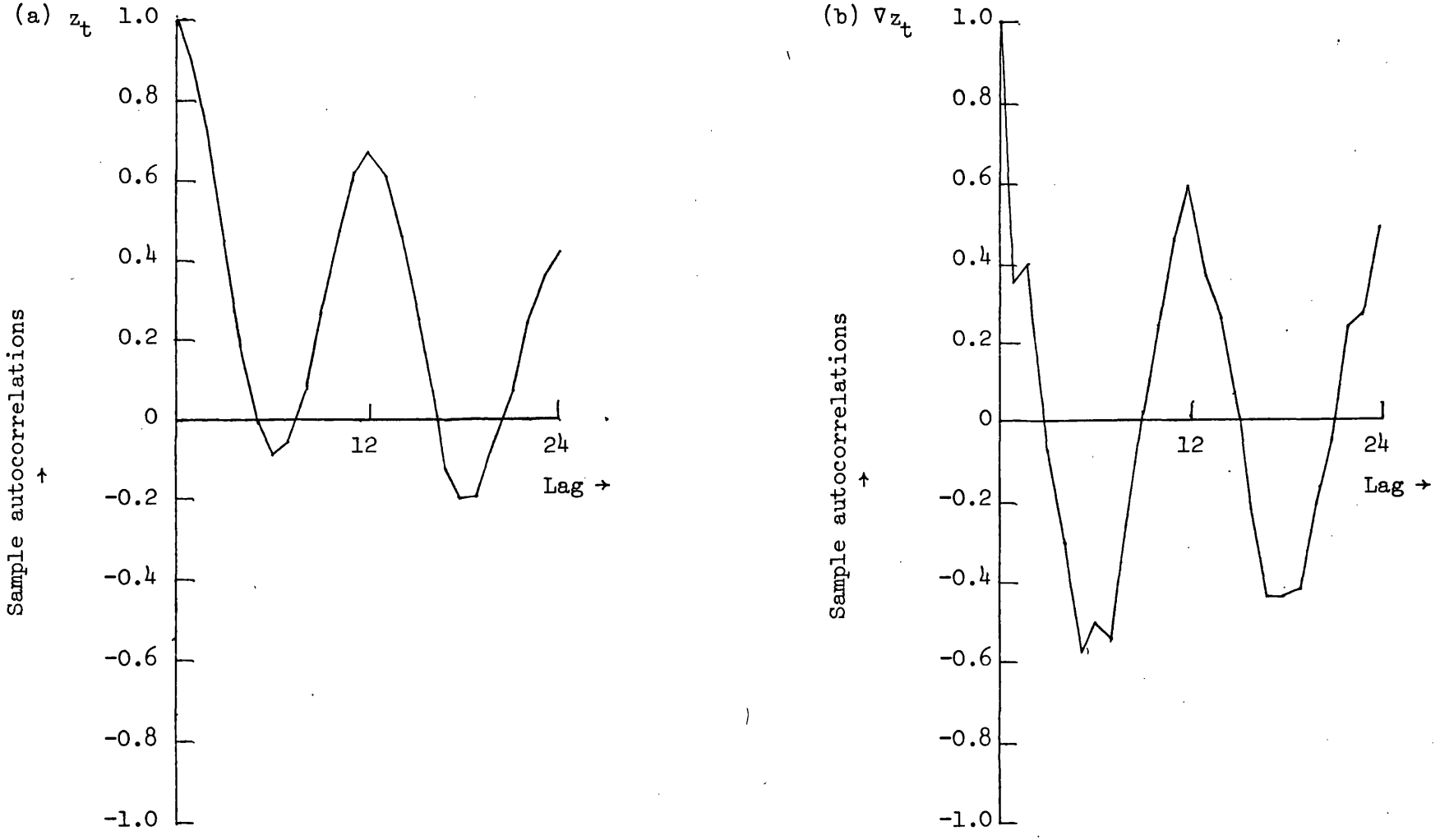
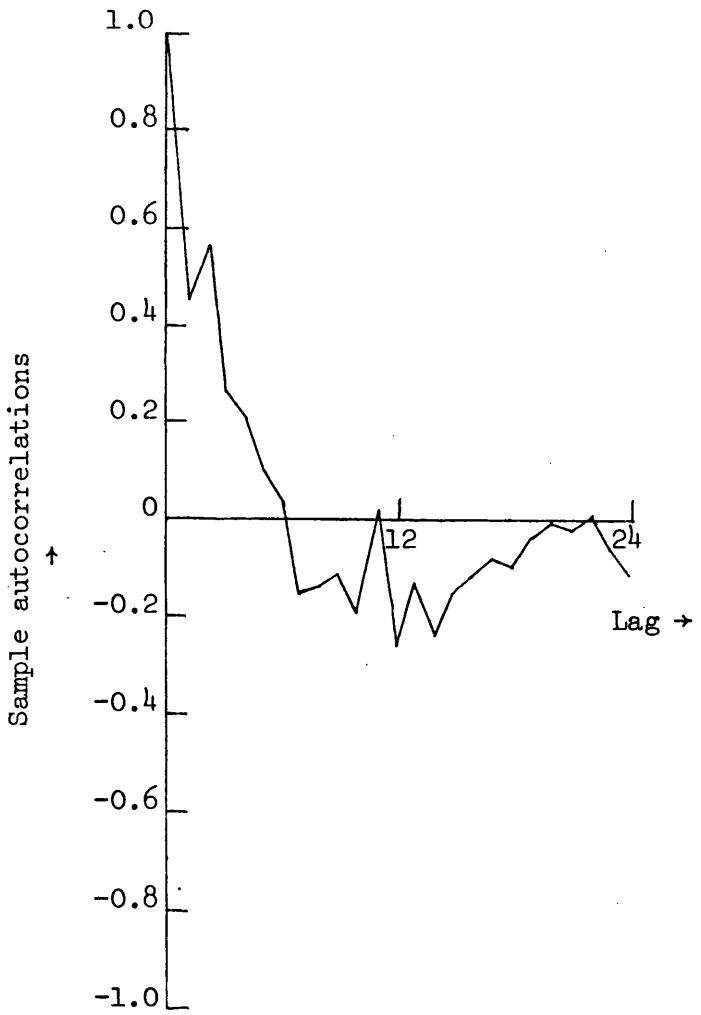


FIGURE 3.4

Sample autocorrelation functions for various differences of the series z_t

(c) $\nabla_{12} z_t$



(d) $\nabla\nabla_{12} z_t$

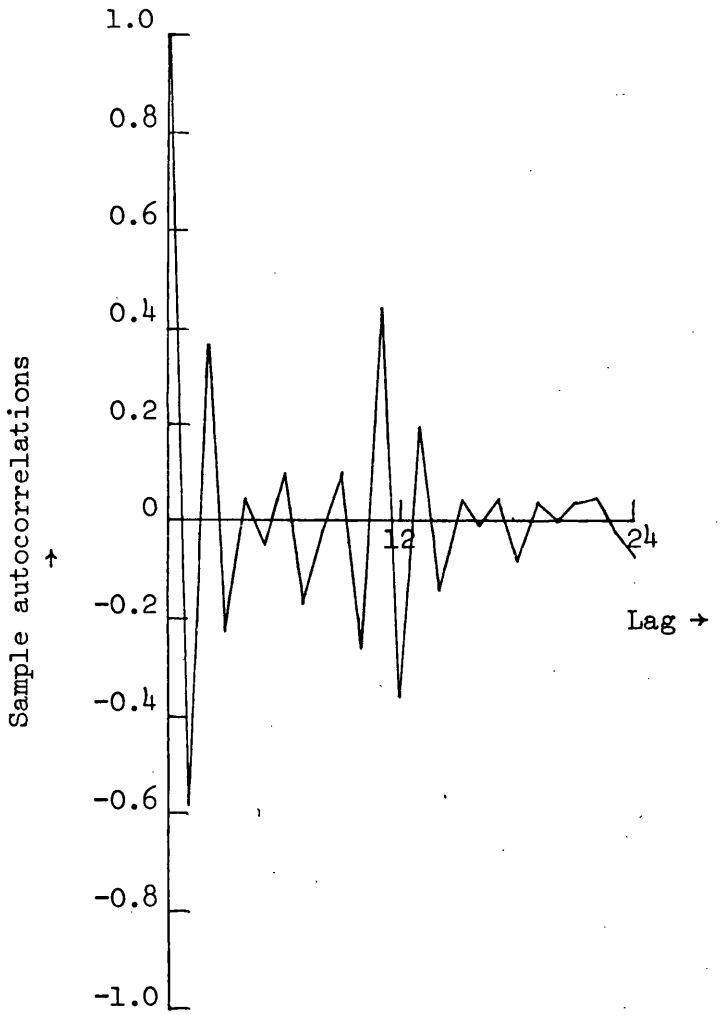


FIGURE 3.4 (cont.)

Sample autocorrelation functions for various differences of the series z_t

It should be noted that the series z_t, ∇z_t, $\nabla_{12} z_t$ and $\nabla\nabla_{12} z_t$ contain 77, 76, 65 and 64 observations respectively. Further differencing was not contemplated since this would reduce the number of terms even more. Values of d = 1 and D = 1 were therefore entertained.

3.3.2 Identifying the Stationary Process

The series $\nabla\nabla_{12} z_t$ is denoted by $w_t (t=1,2,3,\ldots,64)$ i.e.

$$w_t = \nabla\nabla_{12} z_t \qquad 3.3.3$$

the z_t's being defined from t = -12 to t = 64. A seasonal A.R.M.A. model of the form

$$\phi_p(B)\, \Phi_P(B^{12}) w_t = \theta_q(B)\, \Theta_Q(B^{12}) a_t \qquad 3.3.4$$

which will provide a good description of the stationary process w_t must now be identified. Thus values have to be assigned to the integers p, P, q and Q.

As described in Section 2.4.1, reference is made to the sample autocorrelation function of the series w_t. The autocorrelation coefficients at lags 1, 2 and 3 are -0.58, 0.36 and -0.22 and so initially the sample autocorrelation function is decaying by a factor of about -0.6. This suggests the presence of a non-seasonal autoregressive parameter in the model (3.3.4) i.e. p = 1.

On the other hand, the autocorrelation coefficient at lag 12 is "large" while that at lag 24 is "small" and so although no reliable estimates for the sample autocorrelations at lags 36, 48, etc. can be computed, the choice of a seasonal moving average parameter would seem to be reasonable i.e. Q = 1. In the interests of "parsimonious parameterization", P and q were both taken to be zero.

The tentatively identified model was therefore

$$(1 - \phi B)w_t = (1 - \Theta B^{12})a_t \qquad 3.3.5$$

where ϕ and Θ satisfy $|\phi|<1$ and $|\Theta|<1$ in order for the process to be both stationary and invertible.

Using a method similar to that employed by Box and Jenkins (1970, page 74), it can be shown that the theoretical autocorrelation function for the model (3.3.5) is given by

$$\rho(k) = \begin{cases} \phi^k - \dfrac{\phi^{12-k}\Theta(1-\phi^{2k})}{1+\Theta^2-2\phi^{12}\Theta} & \text{for } k = 1,2,3,\ldots,12 \\ \phi\,\rho(k-1) & \text{for } k \geq 13 \end{cases} \qquad 3.3.6$$

Provided ϕ is not too close to ± 1 a good approximation to $\rho(k)$, for certain values of k, is

$$\rho(k) \simeq \begin{cases} \phi^k & \text{for } k = 1,2,3. \\ \dfrac{-\phi^{|12-k|}\Theta}{1+\Theta^2} & \text{for } k = 10,11,12,13,14. \end{cases} \qquad 3.3.7$$

This theoretical autocorrelation function thus compares reasonably favourably with the sample autocorrelation function for w_t, except at lag 11. The sample autocorrelation coefficient at lag 11 is greater than that at lag 12 whereas the theoretical autocorrelation function for model (3.3.5) implies that $\rho(11) < \rho(12)$. The model (3.3.5) would therefore explain some of the high correlation present in w_t at lag 11, but not all of it. A model including an additional moving average parameter at lag 11 may be more successful, but this would take us outside the multiplicative class of seasonal models defined by equation (2.6.5). The extra complications involved in including a further parameter at lag 11 did not seem to be justifiable and the model (3.3.5) was retained.

3.3.3 Preliminary Estimates of Parameters

The final stage of the identification procedure is to obtain preliminary estimates for the model parameters. These estimates are generally used as starting values for the more exact estimation techniques to be described in section 3.4.

Using equation (3.3.7), for $\rho(1)$ and $\rho(12)$, and replacing the theoretical autocorrelation coefficients by their sample estimates we get

$$\left.\begin{aligned}\hat{\phi} &= -0.58\\ \hat{\theta} &= 0.42\end{aligned}\right\} \qquad 3.3.8$$

Again, for model (3.3.5) it can be shown that the variance of the residuals is given by

$$\sigma_a^2 = \frac{(1-\phi^2)}{1+\theta^2-2\phi^{12}\theta}\ \sigma_w^2 \qquad 3.3.9$$

where σ_w^2 is the variance of the stationary series w_t. Thus a preliminary estimate for σ_a^2 can be arrived at by replacing ϕ and θ by their initial estimates and substituting the sample variance of w_t for σ_w^2. This resulted in

$$\hat{\sigma}_a^2 = 0.0063 \qquad 3.3.10$$

3.4 Estimation

Having tentatively identified the model (3.3.5), least squares estimates for the parameters ϕ, θ and σ_a^2 were determined by two methods, the graphical technique outlined in Section 2.4.2 and the non-linear least squares approach described by Box and Jenkins (1970, pages 231-242). Since both methods involved the computation of the unconditional sum of squares, we will begin by illustrating how the latter was evaluated.

3.4.1 Computation of the Unconditional Sum of Squares

In computing the unconditional sum of squares, use is made of the fact that the model (3.3.5) can be expressed in terms of the forward shift operator F, i.e.

$$(1-\phi F)w_t = (1-\theta F^{12})e_t \qquad 3.4.1$$

where the e_t's are independent random variables, normally distributed with zero mean and variance σ_e^2 ($= \sigma_a^2$). Rearranging equations (3.4.1) and (3.3.5) and letting the symbol [ ] denote expectations at time t, conditional on ϕ, θ and $w_1, w_2, w_3, \ldots, w_{64}$ we get

$$[e_t] = [w_t] - \phi[w_{t+1}] + \theta[e_{t+12}] \qquad 3.4.2$$

and

$$[a_t] = [w_t] - \phi[w_{t-1}] + \theta[a_{t-12}] \qquad 3.4.3$$

Setting $[e_{64}], [e_{65}], [e_{66}], \ldots\ldots, [e_{75}]$ equal to zero and $[w_t] = w_t$ (for $t = 1,2,3,\ldots,64$), values for $[e_t]$, for particular values of ϕ and θ, were calculated in reverse order using equation (3.4.2) down to $[e_1]$. Then setting $[e_0], [e_{-1}], [e_{-2}], \ldots\ldots$, equal to zero, equation (3.4.2) was used to back-forecast $[w_0], [w_{-1}], [w_{-2}], \ldots\ldots$, until these values approached zero. For $\phi = -0.60$, $\theta = 0.40$ this point was judged to have occurred at $t = -17$ when $[w_{-17}]$ was less than 0.0005. The $[a_t]$'s for $t = -16,-15,-14,\ldots,64$ were then computed using equation (3.4.3) and setting $[a_{-17}], [a_{-18}], [a_{-19}], \ldots\ldots$, equal to zero. The sum of squares $S(\phi,\theta)$ was calculated using

$$S(\phi,\theta) = \sum_{t=-16}^{64} [a_t]^2$$

A specimen calculation of $S(\phi,\theta)$ for $\phi = -0.60$ and $\theta = 0.40$ is shown in Table 3.3.

Table 3.3 Specimen calculation of $S(\phi,\theta)$, for $\phi = -0.60$, $\theta = 0.40$

z_t	t	$[a_t]$	$[w_t]$	$[e_t]$
	.	.	.	.
	.	.	.	.
	.	.	.	.
	.	.	.	.
	-17	0	0.000	0
	-16	-0.001	-0.001	0
	.	.	.	.
	.	.	.	.
	.	.	.	.
	.	.	.	.
2.188	-12	-0.005	-0.008	0
1.982	-11	0.008	0.013	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
2.389	-1	0.047	0.041	0
2.301	0	-0.006	-0.029	0
2.072	1	-0.038	-0.024	-0.010
1.954	2	-0.016	0.001	0.095
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
2.508	51	0.194	0.227	0.096
2.277	52	-0.073	-0.202	-0.108
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
2.395	63	-0.003	-0.188	-0.025
2.435	64	0.130	0.272	0

$$S(-0.60,0.40) = \sum_{t=-16}^{64} [a_t]^2 = 0.391$$

3.4.2 Graphical Study of the Sum of Squares Function

The sum of squares $S(\phi,\theta)$ was calculated over a grid of values for ϕ and θ using the technique described in Section 3.4.1. In view of the fact that the preliminary estimate for θ, obtained in Section 3.3.3, was not close to ± 1 and also in the light of the statement by Box and Jenkins (1970, page 217) it was decided not to perform more than one iterative cycle for any value of θ. The justification of this decision will be examined in Chapter 6.

The unconditional sum of squares function is tabulated below and the sum of squares surface is illustrated in Figure 3.5.

Table 3.4 Sum of Squares over a grid of values for ϕ and θ

ϕ \ θ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
-0.10	0.601	0.562	0.526	0.492	0.460	0.431	0.403	0.383	0.407	0.703
-0.20	0.543	0.510	0.480	0.452	0.425	0.400	0.376	0.360	0.384	0.664
-0.30	0.497	0.470	0.445	0.421	0.399	0.378	0.358	0.344	0.368	0.635
-0.40	0.463	0.441	0.421	0.401	0.382	0.365	0.348	0.336	0.360	0.616
-0.50	0.442	0.424	0.407	0.391	0.375	0.361	0.347	0.336	0.360	0.608
-0.60	0.433	0.418	0.404	0.391	0.378	0.366	0.354	0.344	0.367	0.610
-0.70	0.436	0.423	0.412	0.401	0.390	0.380	0.369	0.360	0.382	0.623
-0.80	0.451	0.440	0.430	0.421	0.412	0.403	0.392	0.383	0.405	0.649
-0.90	0.478	0.469	0.460	0.451	0.443	0.434	0.424	0.416	0.441	0.710
-1.00	0.518	0.509	0.501	0.492	0.484	0.476	0.468	0.470	0.560	1.211

From Figure 3.5 it can be seen that the sum of squares surface is fairly flat and reasonably quadratic in the neighbourhood of its minimum value. A closer examination of the sum of squares surface for values of ϕ and θ in this neighbourhood revealed that the minimum value occurred when $\phi = -0.45$ and $\theta = 0.81$. Thus we have

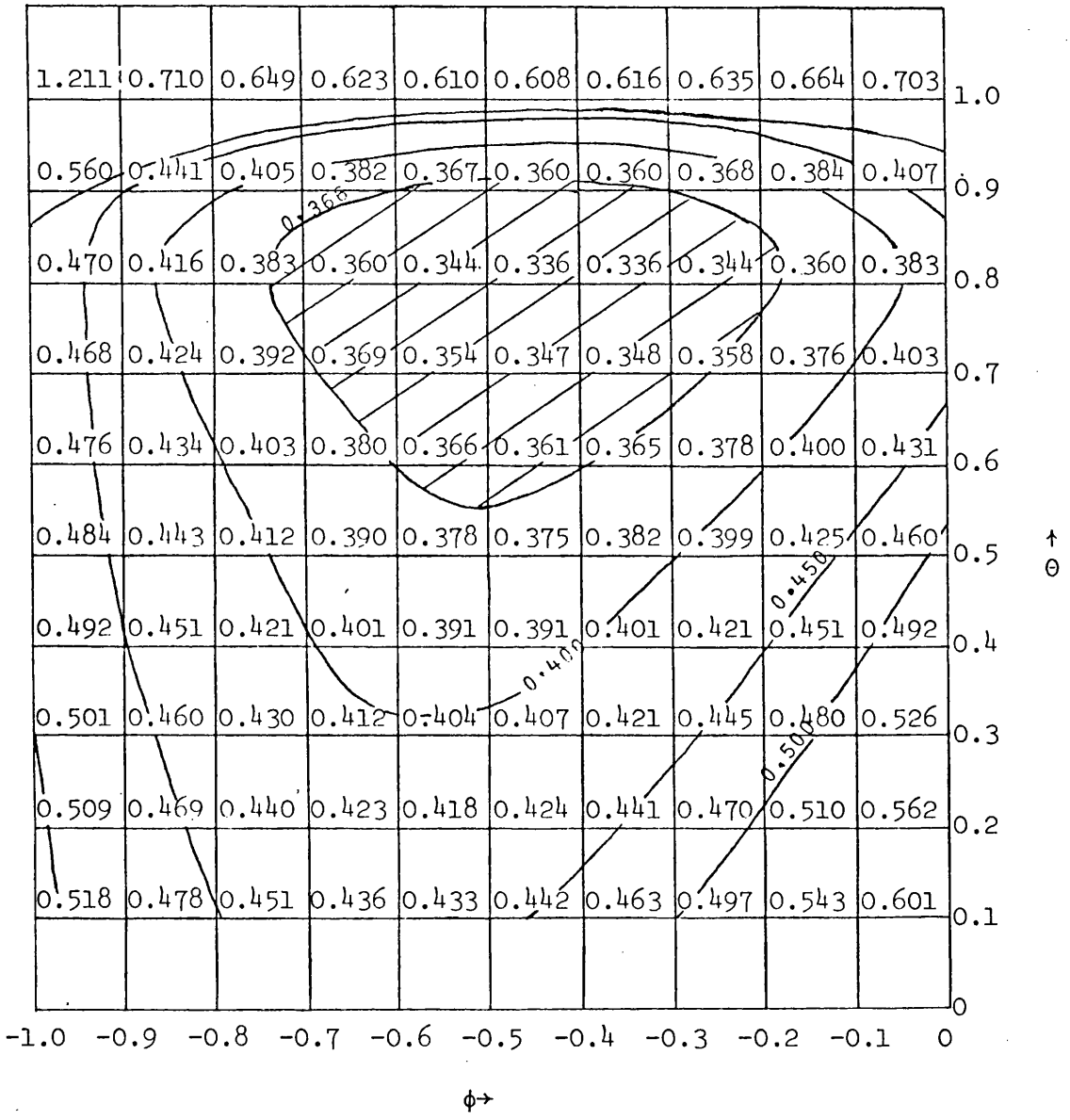


FIGURE 3.5

Sum of squares surface for model $(1-\phi B)w_t = (1-\theta B^{12})a_t$. Shaded area is an approximate 95 per cent confidence region

$$\hat{\phi} = -0.45 \qquad \hat{\theta} = 0.81$$

and figure 3.5 shows that $\hat{\phi}$ and $\hat{\theta}$ are approximately uncorrelated.

A 95% confidence region was constructed following Box and Jenkins (1970, page 229). This is also shown in Figure 3.5.

An estimate of σ_a^2 is given by

$$\hat{\sigma}_a^2 = \frac{S(\hat{\phi},\hat{\theta})}{64}$$

$$= \frac{0.335}{64} = 0.00523$$

(see Box and Jenkins (1970,page 277)).

3.4.3 Non-linear Estimation

The parameters ϕ and θ were also estimated by a non-linear least squares approach using the I.C.L. computer package. Initially ϕ and θ were set equal to zero which differs from the starting values suggested by Box and Jenkins (1970, page 233) who employ the preliminary estimates computed at the identification stage. However the estimates arrived at by the I.C.L. programme were

$$\hat{\phi} = -0.47 \qquad \hat{\theta} = 0.81$$

which agree closely with those arrived at in Section 3.4.2

It is shown in Box and Jenkins (1970, pages 240-242) that for large samples, the variances of $\hat{\phi}$ and $\hat{\theta}$ are given approximately by

$$V(\hat{\phi}) \simeq \frac{1}{64}(1-\phi^2) \text{ and } V(\hat{\theta}) \simeq \frac{1}{64}(1-\theta^2)$$

and hence in our case the standard errors associated with $\hat{\phi}$ and $\hat{\theta}$ are approximately 0.11 and 0.07 respectively.

3.5 Diagnostic Checking

Using the parameter estimates obtained in Section 3.4.3, the fitted model can be expressed as

$$(1+0.47B)w_t = (1-0.81B^{12})a_t \qquad 3.5.1$$

The residuals $\hat{a}_t$ (t=1,2,3,...,64) from the above model are computed by rearranging equation (3.5.1) as

$$\hat{a}_t = w_t + 0.47w_{t-1} + 0.81\hat{a}_{t-12} \qquad 3.5.2$$

where the initial estimates of $\hat{a}_t$ and w_t are derived using the method of back-forecasting described in Section 3.4.1.

The diagnostic checks illustrated in this section are concerned chiefly with the residuals from the fitted model although an example of "overfitting" will be looked at in Section 3.5.3.

3.5.1 Autocorrelations of the Residuals

The sample autocorrelation function of the residuals, for lags 1 to 24, is shown in Table 3.5.

Table 3.5 Sample autocorrelation function of residuals from the model $(1+0.47B)w_t = (1-0.81B^{12})a_t$.

Lags	Autocorrelations											
1-12	0.01	0.10	-0.10	-0.10	-0.16	-0.08	-0.29	-0.06	-0.04	-0.06	0.34	0.08
13-24	0.02	-0.08	-0.02	0.07	-0.04	-0.10	0.04	0.05	0.02	0.04	-0.03	-0.20

Under the assumption that the $\hat{a}_t$'s form a sequence of normally distributed independent random variables, an upper bound for the standard error of the autocorrelations $r_{\hat{a}}(k)$ would be $1/\sqrt{64}$ (Bartlett (1964)). The estimated autocorrelations at lags 7 and 11 both lie outside ±2 standard errors. The "large" autocorrelation at lag 11 is not unexpected since we remarked in Section 3.3.2 that the model

(3.3.5) would not be capable of explaining all of the high correlation present at lag 11 in the series w_t. The autocorrelation coefficient at lag 7 (-0.29) is rather more surprising in view of the fact that the autocorrelation at lag 7 for the series w_t was -0.17 (see Table 3.2). However, the model (3.3.5) cannot really be expected to produce a reduction in correlation from w_t to $\hat{a}_t$, for all lags. Overall it is not too surprising to find 2 "significant" values in 24 coefficients.

The overall adequacy of the fitted model (3.5.1) was tested by the method based on the Q-statistic introduced in Section 2.4.3. After taking into account the amount of data available, this statistic was calculated by summing the squares of the first 36 estimated autocorrelations i.e. K = 36 and

$$Q = 64 \sum_{k=1}^{36} r_{\hat{a}}^2 (k) = 29.07$$

If the fitted model is adequate Q should be distributed approximately as χ^2 with 36 - 1 - 1 = 34 degrees of freedom. The observed value is thus not significant, indicating that the model (3.5.1) is adequate.

3.5.2 Cumulative Periodogram of the Residuals

The normalised cumulative periodogram for the residuals was calculated following the approach outlined in Section 2.4.3. This quantity is plotted in Figure 3.6. Also shown is the theoretical line joining the origin to the point (0.5,1) while Kolmogorov-Smirnov limits are drawn either side of the former. In this case the limit lines were constructed such that for a truly random series they would be crossed 25% of the time. For more information concerning the Kolmogorov-Smirnov limits, the reader is referred to Box and Jenkins (1970, page 297).

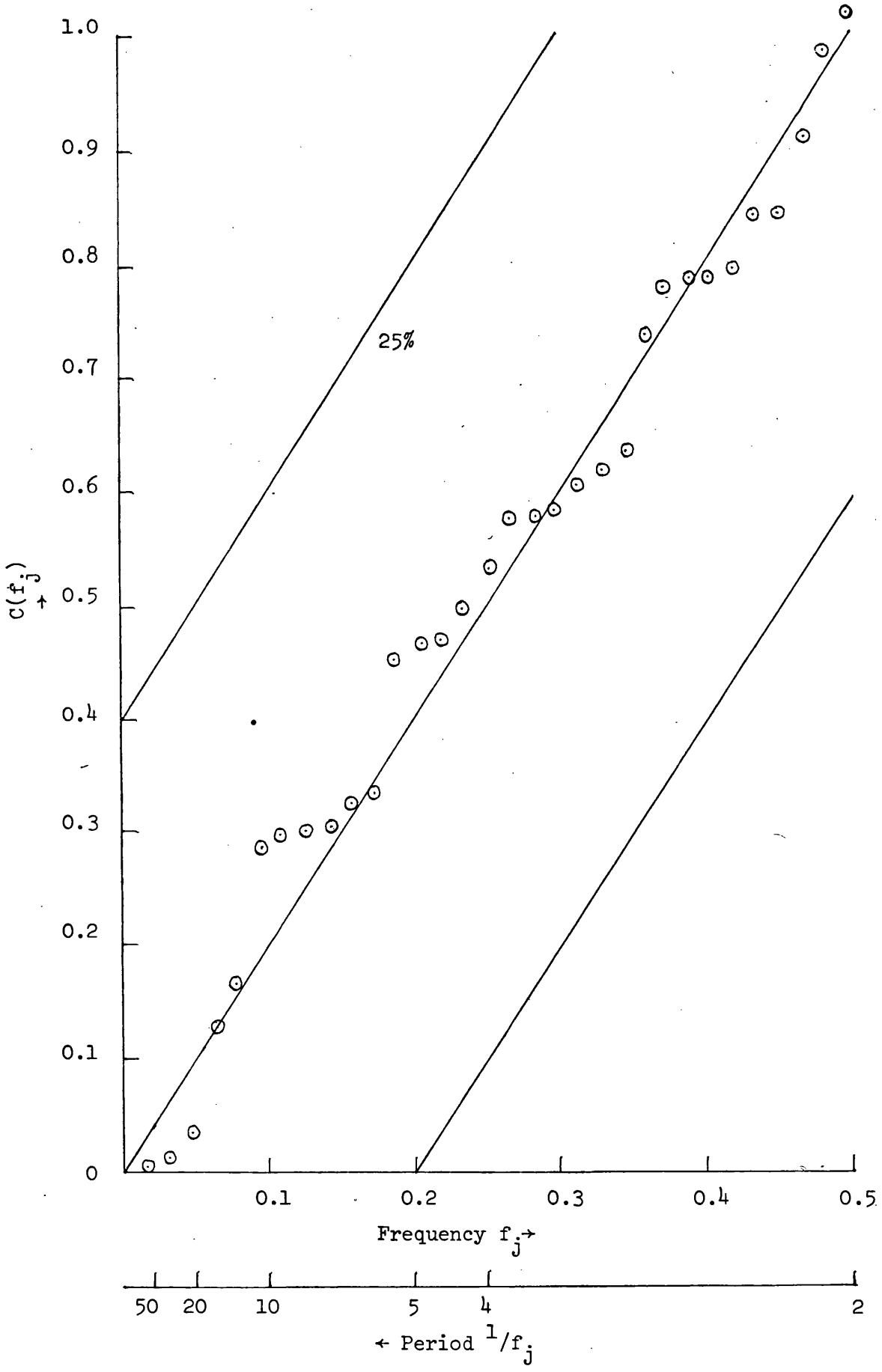


FIGURE 3.6

Normalised cumulative periodogram for the residuals from the model $(1+0.47B)w_t = (1-0.81B^{12})a_t$

It can be seen that the plot of the normalised cumulative periodogram lies well within the 25% limits. However, a slight hump does appear in the function at a frequency between 0.09 and 0.10 corresponding to a period of 10 or 11. It is recalled of course that the autocorrelation of the residuals at lag 11 was also "large". Nevertheless, the cumulative periodogram check does not appear to bring any serious anomalies to light.

3.5.3 Overfitting

The technique of overfitting basically involves the addition of extra parameters to the identified model to cover directions in which discrepancies are most feared (see Box and Jenkins (1970, page 286)).

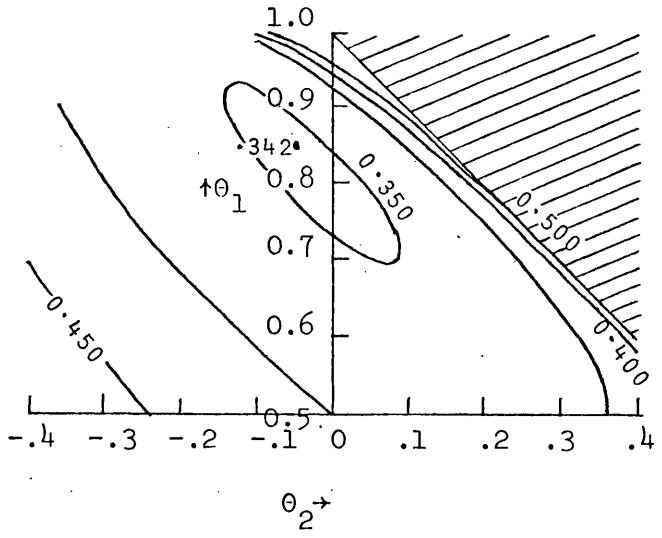
In the case under consideration it was decided to add a second moving average seasonal parameter to the model (3.3.5) in an attempt to explain some of the correlation present in the residuals at lag 24, ($\hat{r_a}(24) = -0.20$). No effort was made to explain the correlation at lag 11, for the reasons given in section (3.3.2). The extended model to be entertained was thus

$$(1-\phi B)w_t = (1-\theta_1 B^{12}-\theta_2 B^{24})a_t \qquad 3.5.3$$

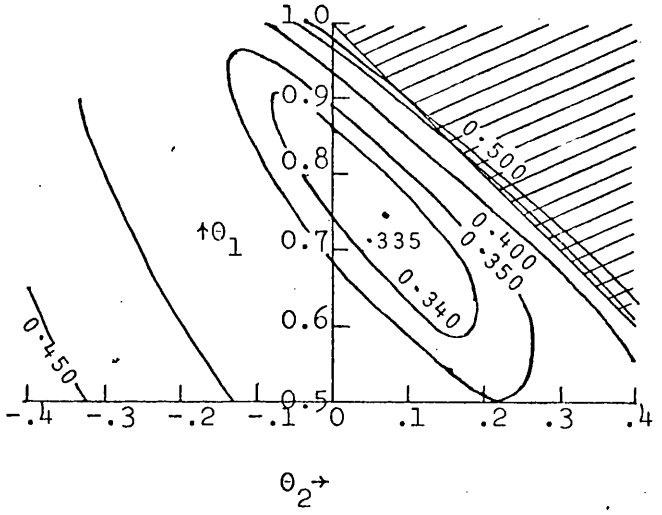
At this stage, the I.C.L. programme produced unsatisfactory estimates for ϕ, θ_1 and θ_2 - estimates which gave rise to a larger residual variance than that possessed by the fitted model (3.5.1). This unfortunate occurrence could probably be attributed to the choice of starting values mentioned in Section 3.4.3.

The problem was overcome by using a graphical technique similar to that employed in Section 3.4.2. The unconditional sum of squares $S(\phi,\theta_1,\theta_2)$ was computed over a grid of values for ϕ, θ_1 and θ_2. Figure 3.7 shows two dimensional contour diagrams

(a) $\phi = -0.30$



(b) $\phi = -0.50$



(c) $\phi = -0.70$

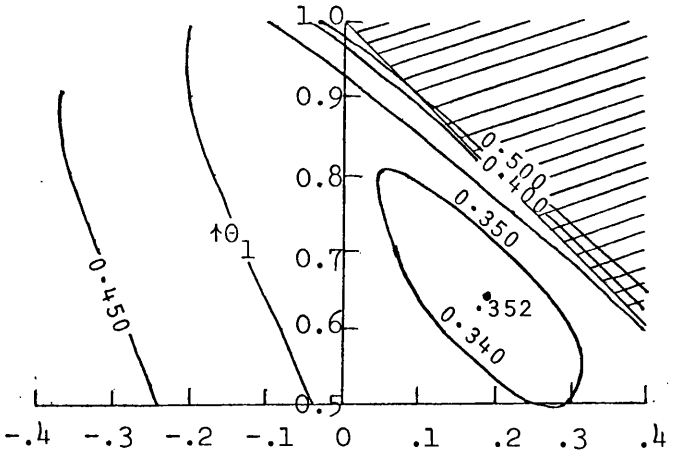


FIGURE 3.7

Sum of squares surface for the model $(1-\phi B)w_t = (1-\theta_1 B^{12}-\theta_2 B^{24})a_t$.
Shaded areas indicate non-invertibility region

for the parameters θ_1 and θ_2, for fixed values of the third parameter ϕ. The fitted model was found to be

$$(1+0.50B)w_t = (1-0.70B^{12}-0.10B^{24})a_t \qquad 3.5.4$$

with

$$\hat{\sigma}_a^2 = 0.00523$$

The minimum value of the sum of squares agrees with that for the two parameter model (3.5.1), up to three decimal places. In fact, Figure 3.7(b) confirms that it is not worthwhile including a third parameter in the model.

In summary, neither the residual analysis nor the overfitting suggested any gross deficiencies in the model (3.5.1). This model was therefore used to generate forecasts.

3.6 Forecasting

Forecasts were evaluated using the difference equation approach described in Section 2.5. The fitted model (3.5.1) is re-written in terms of z_t as

$$z_t = 0.53\, z_{t-1} + 0.47\, z_{t-2} + z_{t-12} - 0.53\, z_{t-13} - 0.47\, z_{t-14} + a_t - 0.81\, a_{t-12} \qquad 3.6.1$$

and following the notation of section 2.5 the forecast made at time t for some future value $z_{t+\ell}$ is given by

$$\hat{z}_t(\ell) = \underset{t}{E}[z_{t+\ell}] = 0.53\, \underset{t}{E}[z_{t+\ell-1}] + 0.47\, \underset{t}{E}[z_{t+\ell-2}] + \underset{t}{E}[z_{t+\ell-12}] - 0.53\, \underset{t}{E}[z_{t+\ell-13}] - 0.47\, \underset{t}{E}[z_{t+\ell-14}] + \underset{t}{E}[a_{t+\ell}] - 0.81\, \underset{t}{E}[a_{t+\ell-12}] \qquad 3.6.2$$

with

$$\text{i)}\quad {}_{t}E[z_{t+j}] = \hat{z}_t(j)\ ,\ {}_{t}E[a_{t+j}] = 0 \text{ for } j = 1,2,3,\ldots$$

$$\text{ii)}\quad {}_{t}E[z_{t-j}] = z_{t-j}\ ,\ {}_{t}E[a_{t-j}] = a_{t-j} = z_{t-j} - \hat{z}_{t-j-1}(1)$$

$$\text{for } j = 0,1,2,\ldots$$

As explained in Section 2.5 the accuracy of the point forecasts derived from equation (3.6.2) can be assessed by expressing the model (3.5.1) as an infinite moving average

$$z_t = (1+\psi_1 B + \psi_2 B^2 + \ldots\ldots)a_t \qquad 3.6.3$$

where the ψ_j's are arrived at by equating the coefficients of various powers of B in the equation

$$(1+0.47B)\ (1-B)\ (1-B^{12})\ (1+\psi_1 B+\psi_2 B^2+\ldots\ldots) = (1-0.81B^{12})$$

These ψ weights are quoted in Table 3.6.

Table 3.6 The ψ weights for the process $(1+0.47B)w_t = (1-0.81B^{12})a_t$

j	1	2	3	4	5	6	7	8	9	10	11	12
ψ_j	0.54	0.75	0.65	0.70	0.67	0.68	0.68	0.68	0.68	0.68	0.68	0.87

The ψ weights are used in computing the variance of the forecast errors, for each lead time, following the theory of Section 2.5.

Forecasts made at May 1971* for lead times 1 to 12 are illustrated in Figure 3.8. These point forecasts together with associated tolerance limits are given in Table 3.7. The tolerance limits are taken to be ±2x (estimated standard deviation of the forecast errors).

---

* Recalling that the series z_t is defined from t = -12 to 64, May 1971 corresponds to t = 64.

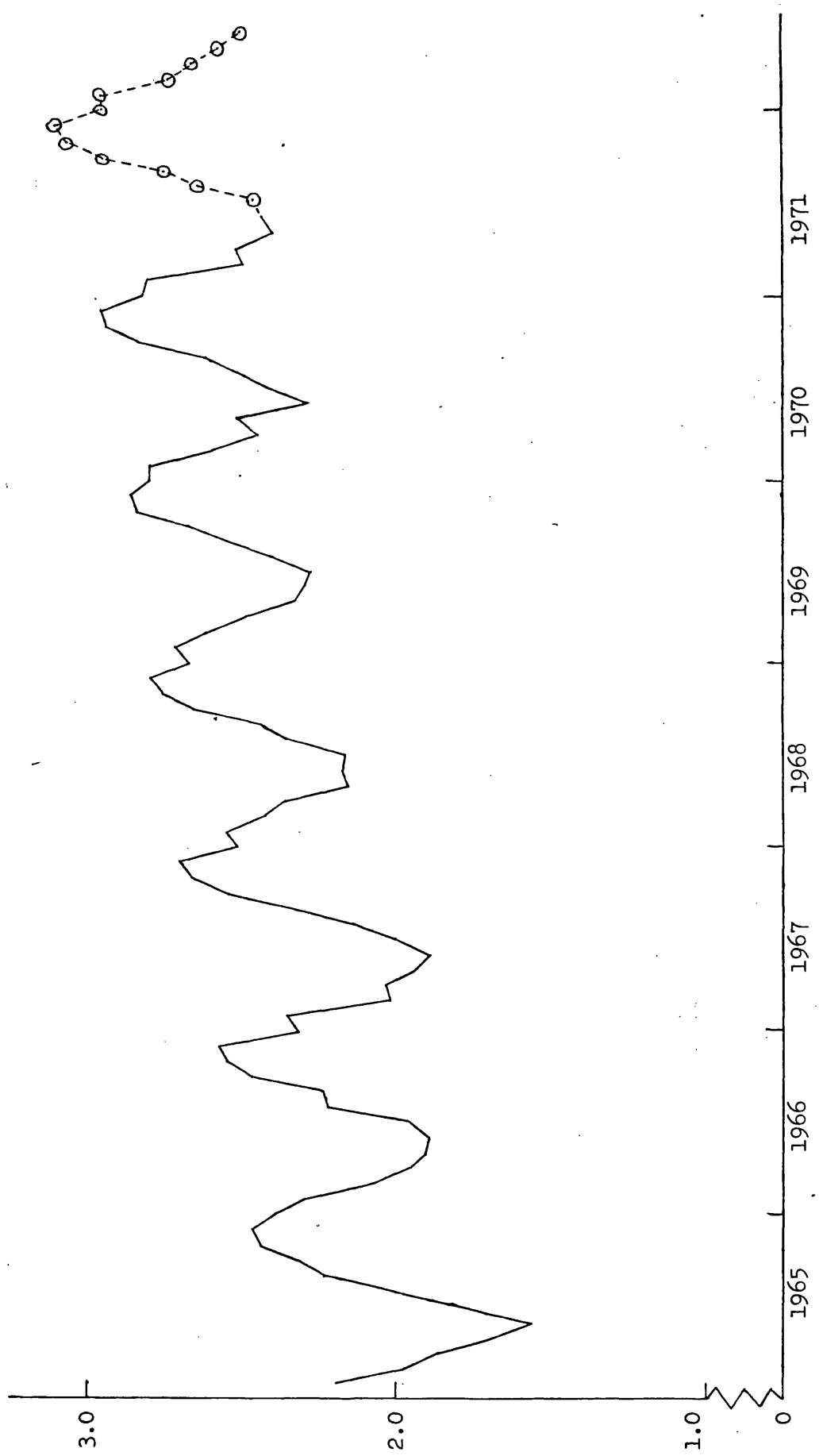


FIGURE 3.8

Forecasts made at May 1971, for lead times 1 - 12

Table 3.7 Forecasts (with tolerance limits) made at May 1971 for lead times 1 to 12

ℓ	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{z}_{64}(\ell)$	2.456	2.641	2.750	2.945	3.060	3.087	2.953	2.949	2.728	2.655	2.565	2.497
T.L.	±0.145	±0.164	±0.196	±0.218	±0.240	±0.259	±0.277	±0.294	±0.310	±0.325	±0.340	±0.354

The forecasts shown in Figure 3.8 and Table 3.7 are of course in terms of the transformed variable z_t. More important are the forecasts in terms of the original variable X_t. Since $z_t = \log_{10} X_t$, the simplest forecast of $X_{t+\ell}$ made at time t is given by

$$\hat{X}_t(\ell) = 10^{\hat{z}_t(\ell)} \qquad 3.6.4$$

Although the forecast errors in terms of the z_t's are assumed to have a zero mean, the forecast errors in terms of the X_t's will not have a zero mean. Granger and Newbold (1970) say that a "bias" is introduced by the transformation and that the forecast defined by equation (3.6.4) is not optimal. Further, the same authors show that if the errors $e_{t+\ell} = z_{t+\ell} - \hat{z}_t(\ell)$ are normally distributed with zero mean and variance σ^2_ℓ then the percentage bias in the forecast $10^{\hat{z}_t(\ell)}$ is $\{\exp[\frac{\sigma^2_\ell}{2}(\log_e 10)^2] - 1\} \times 100\%$. For "small" values of σ^2_ℓ, the latter expression can be adequately replaced by the linear approximation $\{\frac{\sigma^2_\ell}{2}(\log_e 10)^2\} \times 100\%$. In our case estimates for σ^2_ℓ were less than 0.04 for k = 1,2,3,...,12 and so the approximate formula was employed. The percentage bias was generally quite small and although it did become more pronounced as the lead time increased (about 8% for lead time 12), the re-transformed forecasts were computed using equation (3.6.4). These forecasts together with approximate tolerance limits are given in Table 3.8 and plotted in Figure 3.9.

Table 3.8 Forecasts (with approximate tolerance limits) of $X_{t+\ell}$ made at May 1971 for lead times 1-12

ℓ	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{X}_{64}(\ell)$	286	437	562	881	1148	1221	897	889	535	452	367	314
T.L.	±103	±186	±315	±521	±759	±899	±721	±735	±471	±412	±336	±300

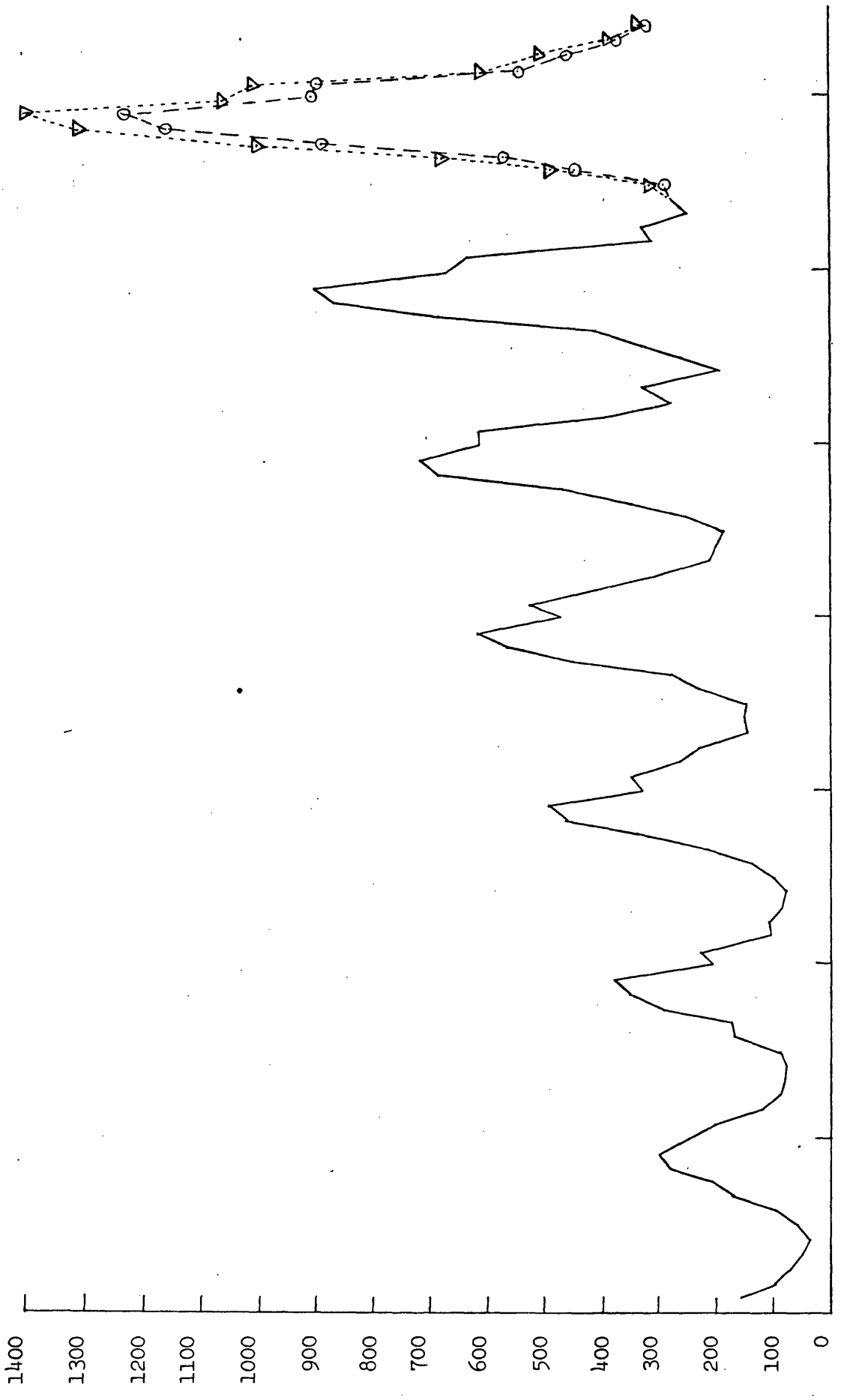


FIGURE 3.9

Forecasts made at May 1971, for lead times 1 - 12.

Initial forecasts ∀----∀----∀; forecasts using back forecasted residuals O--O--O

At this stage it should be pointed out that the forecasts given in Table 3.7 (and hence those in Table 3.8) were computed using the fitted model (3.5.1) with

$$E_t[a_{52+\ell}] = \hat{a}_{52+\ell}, \text{ for } \ell = 1,2,3,\ldots,12$$

$\hat{a}_{52+\ell}$ denoting the residual obtained by the method of back-forecasting. As described in Chatfield and Prothero (1973 a), the I.C.L. programme generated forecasts by setting

$$E_t[a_{52+\ell}] = \hat{a}_{52+\ell}(\text{zero}), \text{ for } \ell = 1,2,3,\ldots,12$$

where $\hat{a}_{52+\ell}(\text{zero})$ represents the residual at time $52 + \ell$, initial unknown residuals being taken to be zero as suggested by Box and Jenkins (1970, page 131). Using equation (3.5.2) a relationship can be established between the residuals $\hat{a}_t$ and $\hat{a}_t(\text{zero})$. For example when $t = 58$ we find that

$$\hat{a}_{58} = \hat{a}_{58}(\text{zero}) + (0.81)^5 \hat{a}_{-2} \qquad 3.6.5$$

and thus a quite considerable amount of weight is given to the back-forecasted residual $\hat{a}_{-2}$. If the latter is non-zero then a substantial difference may exist between $\hat{a}_{58}$ and $\hat{a}_{58}(\text{zero})$. In actual fact, in terms of the transformed observations, the forecasts computed using the residuals $\hat{a}_t(\text{zero})$ did not seem to differ greatly from those generated using the back-forecasted residuals. However, Figure 3.9 shows that in terms of the original observations the difference between the two sets of forecasts is quite marked, particularly in the peaks of the data. It would therefore seem to be reasonable to recommend the use of back-forecasted residuals for forecasting purposes, especially when a transformation is involved.

All future forecasts computed in this thesis will be based on back-forecasted residuals.

Visual inspection of the forecasts shown in Figure 3.9 suggests that these forecasts are much higher, particularly near the peak, then one could reasonably have expected. Since the immediate requirement of Company X was forecasts for lead times up to 12 months from May 1971, it was decided to seek some plausible alternatives to the fitted model (3.5.1).

3.7 Some Alternative Models

Although the autocorrelation function of the series w_t led naturally to model (3.3.5), we have seen that the forecasts made for the lead times specifically required were intuitively very poor. An attempt was therefore made to find an alternative model which possessed a similar autocorrelation function to the model (3.3.5) but which produced more reasonable forecasts from May 1971.

Restricting ourselves to two-parameter models, there are four possible models capable of accounting for high autocorrelations at lags 1, 2, 11 and 12. These models are

$$(1 - \phi B)w_t = (1 - \theta B^{12})a_t \qquad (A)$$

$$(1 - \phi B)\,(1 - \Phi B^{12})w_t = a_t \qquad (B)$$

$$(1 - \Phi B^{12})w_t = (1 - \theta B)a_t \qquad (C)$$

and

$$w_t = (1 - \theta B)\,(1 - \theta B^{12})a_t \qquad (D)$$

where in all cases $w_t = \nabla \nabla_{12} \log_{10} X_t$.

Hereafter these models will be referred to as model (A), model (B), etc. Model (A) is of course the initially identified model.

The four models were compared from both a "fitting" point of view and a forecasting point of view. The results of these comparisons will now be discussed.

3.7.1 Fitting the Models to the Data

The models (B), (C) and (D) were fitted to the whole data by exactly the same procedure as that described for model (A). The parameters in each model were estimated by the two techniques employed in Section 3.4. Again there was a good agreement between the two different estimates. Using the estimates derived by the non-linear least squares approach, the four fitted models were

$$(1 + 0.47B)w_t = (1 - 0.81B^{12})a_t$$

$$(1 + 0.51B)\ (1 + 0.47B^{12})w_t = a_t$$

$$(1 + 0.56B^{12})w_t = (1 - 0.49B)a_t$$

and $$w_t = (1 - 0.44B)\ (1 - 0.85B^{12})a_t$$

while the estimated variances of the residuals ($\hat{\sigma}_a^2$) for each model are shown in Table 3.9 .

Table 3.9 Estimates of residual variance

Model	$\hat{\sigma}_a^2$
(A)	0.00523
(B)	0.00659
(C)	0.00694
(D)	0.00539

The residual variance is smallest for model (A), indicating that this model fits the data better than the 3 alternatives. However, it seems doubtful whether the differences between the four estimates are statistically or practically significant - certainly there appears to be no "significant" difference between the estimated variances of (A) and (D) or between (B) and (C). The problem of comparing the fit of

two models including the same number of parameters has not been discussed by Box and Jenkins (1970) although tests have been suggested by Whittle (1952) and Walker (1967).

The residuals were estimated by the method of back-forecasting and a visual inspection emphasised the similarities between models (A) and (D) and between models (B) and (C).

An examination of the autocorrelation functions of the residuals for each model revealed that "large" autocorrelations at lags 7 and 11 were present in all four cases. In addition, other significant autocorrelation coefficients occurred at lag 24 for models (B) and (C) and at lag 2 for model (D). The presence of these autocorrelations in the models (B), (C) and (D) seemed to justify the use of a model including a non-seasonal autoregressive parameter and a seasonal moving average parameter i.e. model (A).

As an overall test of the adequacy of the four models the Q-statistic was calculated in each case and the results are shown in Table 3.10. The squared estimated autocorrelations were summed over lags 1 to 36.

Table 3.10 Values of Q

Model	Q
(A)	29.07
(B)	38.60
(C)	35.87
(D)	33.94

Model (A) produced the smallest value for Q suggesting that the assumptions made about the residuals were rather more valid for this model than for models (B), (C) and (D). Even so, none of the Q-values were significantly large when compared with the percentage points of the χ^2 distribution with 34 degrees of freedom. Thus it was concluded

that no serious inadequacies were present in any of the four models.

3.7.2 Forecasting

Point forecasts were computed, using all four models, from May 1971 for lead times 1 to 12. The forecasts resulting from models (A) and (D) and from models (B) and (C) were so similar that only those from model (A) and model (B) are shown in Figure 3.10. From a purely intuitive standpoint, the model (B) point forecasts look far more reasonable than those of model (A). Although the former are more satisfactory than the latter, the tolerance limits associated with the model (B) forecasts were found to be much wider than expected. For example, the tolerance limits for lead time 3 are 452 ± 227.

An interesting point concerning the two sets of forecasts shown in Figure 3.10 is that the difference between the point forecasts from the 2 models is of a much higher order than that which can occur through errors in estimating the parameters within each model. As an example, the forecast for November 1971 resulting from model (B) is 1,025 when $\phi = -0.80$ and $\Phi = -0.20$ and 978 when $\phi = -0.70$ and $\Phi = -0.50$ as compared with 992 using the estimated parameters $\hat{\phi} = -0.47$ and $\Phi = -0.51$. On the other hand, the corresponding forecast generated by model (A) is 1221.

Thus far only a simple visual comparison between the forecasts made from just one origin has been attempted. A more general comparison of the forecasting performance of each of the models was also undertaken.

Each of the four ~~models were~~ models was fitted to the first 60 observations of the logged data z_t, the parameters being estimated by the non-linear least squares technique. The resulting fitted models were

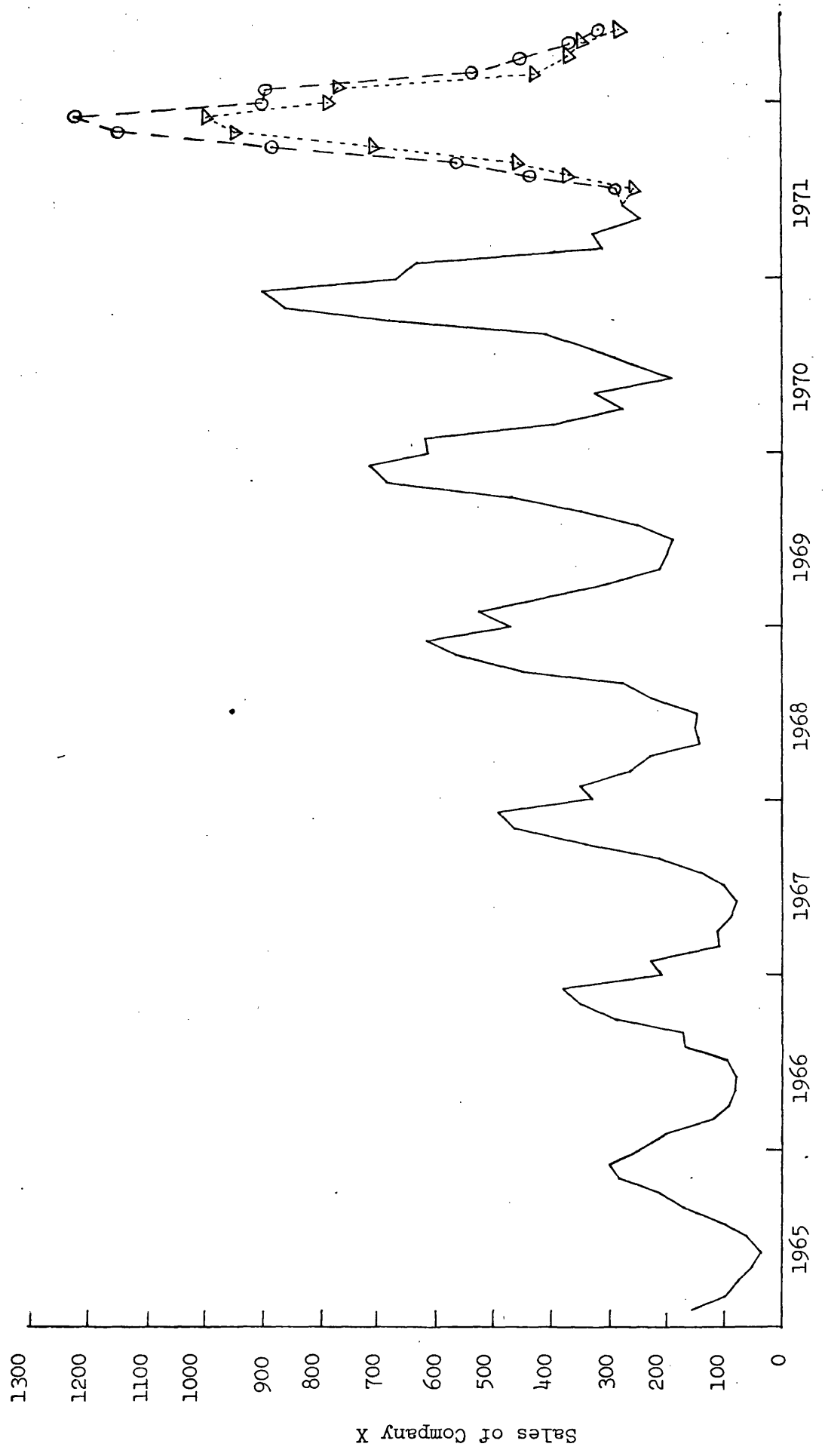


FIGURE 3.10

Revised forecasts from May 1971, using model A, and forecasts using model B.

Model A 0--0--0; model B ∇----∇----∇

$$(1 + 0.37B)w_t = (1 - 0.80B^{12})a_t$$

$$(1 + 0.27B)\ (1 + 0.62B^{12})w_t = a_t$$

$$(1 + 0.61B^{12})w_t = (1 - 0.28B)a_t$$

and $$w_t = (1 - 0.29B)\ (1 - 0.79B^{12})a_t$$

The forecasting potential of the four models was assessed over the remaining 17 observations. The forecasts (in terms of the z_t's) were compared using the mean squared error function M.S.E(ℓ), where ℓ refers to the lead time. In our case the mean squared error was defined by

$$M.S.E(\ell) = \frac{1}{17-\ell+1} \sum_{t=47+\ell}^{64} (z_t - \hat{z}_{t-\ell}(\ell))^2$$

Table 3.11 gives the above function for each model and for lead times 1 to 12.

Table 3.11 Mean squared errors, lead times 1-12

Lead Time	Model (A)	Model (B)	Model (C)	Model (D)
1	0.0060	0.0105	0.0107	0.0067
2	0.0058	0.0096	0.0091	0.0056
3	0.0081	0.0141	0.0134	0.0083
4	0.0075	0.0124	0.0120	0.0077
5	0.0053	0.0118	0.0114	0.0058
6	0.0065	0.0117	0.0114	0.0069
7	0.0087	0.0137	0.0132	0.0091
8	0.0103	0.0146	0.0142	0.0111
9	0.0129	0.0143	0.0135	0.0131
10	0.0167	0.0202	0.0193	0.0176
11	0.0134	0.0127	0.0124	0.0141
12	0.0178	0.0229	0.0221	0.0189

It can be seen from Table 3.11 that model (A) gave rise to the smallest mean squared errors for all lead times except 2 and 11 when models (4) and (3) respectively were best. On account of the small sample over which the mean squared errors were computed (e.g. 6 observations for lead time 12) it would be unwise to attach too much importance to these results. Nevertheless the similarities between models (A) and (D) and between models (B) and (C), noted in section 3.7.1, are again apparent.

As well as obtaining some quantitative measure of each model's forecasting ability, a visual inspection of the individual forecasts over the final 17 observations was also carried out. Models (A) and (D) were found to yield very good forecasts from December 1969, for all lead times, while the forecasts made from May 1970 were generally very poor. On the other hand, models (B) and (C) gave rather more consistent forecasts and seemed less dependent on the month from which the forecasts were being made.

3.8 Conclusions

The results obtained from the Box-Jenkins analysis described in this chapter were unfavourable. The main disappointment was the failure to achieve the original objective, namely that of finding a model capable of producing a satisfactory set of forecasts from May 1971. The initially identified model (A) generated point forecasts which were subjectively far too high and an alternative model was sought. Reasonable point forecasts were provided by the model (B) (and (C)) but the tolerance limits associated with these forecasts were exceptionally wide, especially for the higher lead times. A closer examination of the forecasting performance of the models (A) and (B) over the final 17 observations revealed that the magnitude of the forecast errors resulting from the former tended to depend on the month from which the forecasts were being made while for model (B) this was not so. All the evidence suggested

that model (B) would be a better proposition than model (A) although the fact that its selection would be on subjective grounds, rather than via the identification procedure recommended by Box and Jenkins (1970), was in itself unsatisfactory.

Obviously the most immediate problem is to seek reasons why the identified model (A) failed to generate a reasonable set of forecasts from May 1971 and also why more accurate forecasts were generally derived from base points in the peak of the seasonal cycle rather than in the troughs. As mentioned in the conclusions of Chatfield and Prothero (1973 a) we suspected that the use of the logarithmic transformation may have been at the root of these problems. For this reason, Chapter 4 will be devoted to the subject of transformations with particular reference to the analysis described in this chapter.

Meanwhile, a number of interesting points arose during this case-study, regardless of the transformation employed.

The estimated autocorrelation function of the series $w_t = \nabla \nabla_{12} z_t$ exhibited a "large" value at lag 11. This was rather unexpected and may have arisen to a certain extent through use of the differencing operator $\nabla \nabla_{12}$. The possibility of autocorrelations being induced by the differencing operation is one of the points looked at in Chapter 5 which deals generally with the concept of differencing as a means of producing stationarity.

A rather disturbing feature of the analysis of the Company X data was that although the diagnostic checks did not reveal anything seriously wrong with the models (A), (B), (C) or (D), the models (A) and (D) produced point forecasts from May 1971 which differed considerably from those generated by models (B) and (C). All of these models were based on the differencing operator $\nabla \nabla_{12}$. In Section 5.5 we shall see that a model based on the differencing operator ∇_{12} will fit the data equally as well as model (A), yet the tolerance limits associated with the two models become quite different as the lead time increases.

The similarities between models (A) and (D) and between models (B) and (C) were noted throughout Section 3.7. On reflection, the common parameter in models (A) and (D) was the moving average seasonal parameter while the autoregressive seasonal parameter was common to models (B) and (C). Thus the behaviour of each model tended to be determined by the seasonal parameter and the choice of a moving average or autoregressive non-seasonal parameter was relatively unimportant.

One further feature of the Box-Jenkins procedure apparent from the analysis described in this chapter is the fact that it is not obvious what autoregressive-moving average models actually tell us about the data in terms of the more familiar concepts of trend and seasonality. This point will be expanded in Chapter 9.

CHAPTER 4

TRANSFORMATIONS

4.1 Introduction

The Box-Jenkins forecasting procedure generates forecasts which are a linear function of current and past values of the variable being forecasted, i.e. it is a linear technique. However, many time series encountered in practice possess non-linear properties, e.g. monthly sales data often exhibit a multiplicative seasonal variation. When a non-linear model is appropriate, Box and Jenkins (1970, page 94) suggest transforming the data into a form to which a linear model may reasonably be fitted. Forecasts for future values of the transformed variable are computed using the techniques described in Chapter 2 and these are then transformed back in terms of the original variable.

The data analysed in Chapter 3 displayed a multiplicative seasonal effect and so in order to employ the Box-Jenkins procedure a non-linear transformation was first required. For reasons given in Section 3.2 and also in Chatfield and Prothero (1973a) a logarithmic transformation was applied to the original series. In the discussion following the Chatfield-Prothero paper, Dr. G. Tunnicliffe Wilson suggested that this choice of transformation was the cause of some of the problems raised by the paper. In particular Dr. Wilson demonstrated that more satisfactory forecasts could be obtained by allowing for the wider range of transformations considered by Box and Cox (1964).

This chapter deals with the subject of transformations as applied in conjunction with the Box-Jenkins forecasting procedure. In Section 4.2 the effect of making a logarithmic transformation,

when confronted with data of a form similar to that analysed in Chapter 3, will be examined. The results obtained in Chapter 3 based on a logarithmic transformation will be compared (in Section 4.3) with those arrived at by Dr. Wilson, using a different non-linear transformation. Section 4.4 will contain comments on some of the problems encountered when applying non-linear transformations, in particular the class proposed by Box and Cox (1964), to time series. General comments and conclusions will be given in Section 4.5.

4.2 The Effect of a Logarithmic Transformation

In Section 3.2 it was observed that the sales of Company X possessed a roughly linear trend together with an approximate multiplicative seasonal variation. In order to examine the validity of employing a logarithmic transformation in such an instance, it will be assumed that a series X_t is composed of a purely deterministic linear trend and a multiplicative seasonal pattern with period 12. Thus we may write

$$X_t = (\alpha + \beta t)\, s_t \qquad 4.2.1$$

where α, β and s_j ($j = 1, 2, 2, \ldots, 12$) are constants, the s_j's representing the seasonal effect for each period and $s_t = s_{t-12}$.

Alternatively, equation (4.2.1) may be re-written as

$$X_{t+u} = (\alpha(t) + \beta(t)\, u)\, s_{t+u} \qquad 4.2.2$$

for $u = 0, \pm 1, \pm 2, \ldots$ etc.

where $\alpha(t)$, $\beta(t)$ represent respectively the level of the series and the slope at time t. In terms of equation (4.2.1) $\alpha(t) = \alpha + \beta t$ (i.e. $\alpha(o) = \alpha$), and $\beta(t) = \beta$, for all t, so that equation (4.2.2)

becomes

$$X_{t+u} = (\alpha(t) + \beta u)\, s_{t+u} \text{ , for } u = 0, \pm 1, \pm 2, \ldots \text{ etc} \qquad 4.2.3$$

Now performing a logarithmic transformation on X_{t+u} we get

$$z_{t+u} = \log X_{t+u} = \log [(\alpha(t) + \beta u)\, s_{t+u}]$$

$$= \log (\alpha(t) + \beta u) + \log s_{t+u}$$

$$= \log \alpha(t) + \log\left[1 + \frac{\beta}{\alpha(t)} u\right] + s'_{t+u}$$

$$= A(t) + \log\left[1 + \frac{\beta}{\alpha(t)} u\right] + s'_{t+u} \qquad 4.2.4$$

where $s'_{t+u} = \log s_{t+u} = \log s_{t+u-12} = s'_{t+u-12}$, and $A(t) = \log \alpha(t)$.

Although equation (4.2.4) now represents an additive model at time t, the trend component is no longer linear in u. In fact, for $|u| < \left|\frac{\alpha(t)}{\beta}\right|$ we have

$$\log\left[1 + \frac{\beta}{\alpha(t)} u\right] = \frac{\beta}{\alpha(t)} u - \left[\frac{\beta}{\alpha(t)}\right]^2 u^2 + \left[\frac{\beta}{\alpha(t)}\right]^3 u^3 \ldots\ldots\ldots \qquad 4.2.5$$

which will be approximately linear for smaller values of u only if the ratio $\beta/\alpha(t)$ is also small, i.e. if the monthly growth rate is small. For the sales of Company X, approximate values for $\alpha(o)$ and β were 100 and 5 respectively, representing an initial monthly growth rate of about 5%.

Figure 4.1(a) shows a purely deterministic series constructed with $\alpha = 100$, $\beta = 5$, $s_1 = 1.2$, $s_2 = 0.8$, $s_3 = 0.6$, $s_4 = 0.4$, $s_5 = 0.4$, $s_6 = 0.4$, $s_7 = 0.8$, $s_8 = 1.0$, $s_9 = 1.4$, $s_{10} = 1.8$, $s_{11} = 1.8$ and $s_{12} = 1.4$. The logarithmic transformation of this series is plotted

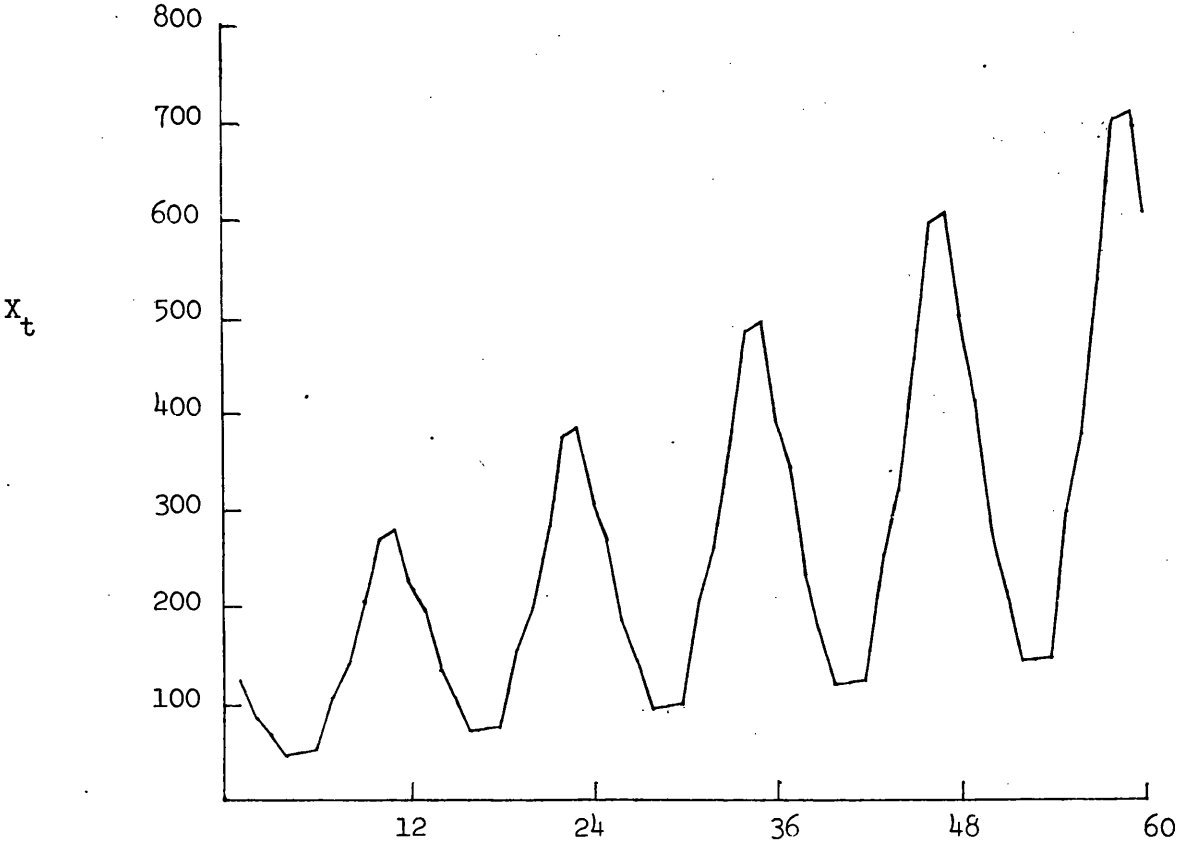


FIGURE 4.1(a)

A purely deterministic series given by

$X_t = (\alpha+\beta t)s_t$, where $\alpha=100$, $\beta=5$, $s_1=1.2$, $s_2=0.8$, $s_3=0.6$, $s_4=0.4$, $s_5=0.4$, $s_6=0.4$, $s_7=0.8$, $s_8=1.0$, $s_9=1.4$, $s_{10}=1.8$, $s_{11}=1.8$, $s_{12}=1.4$.

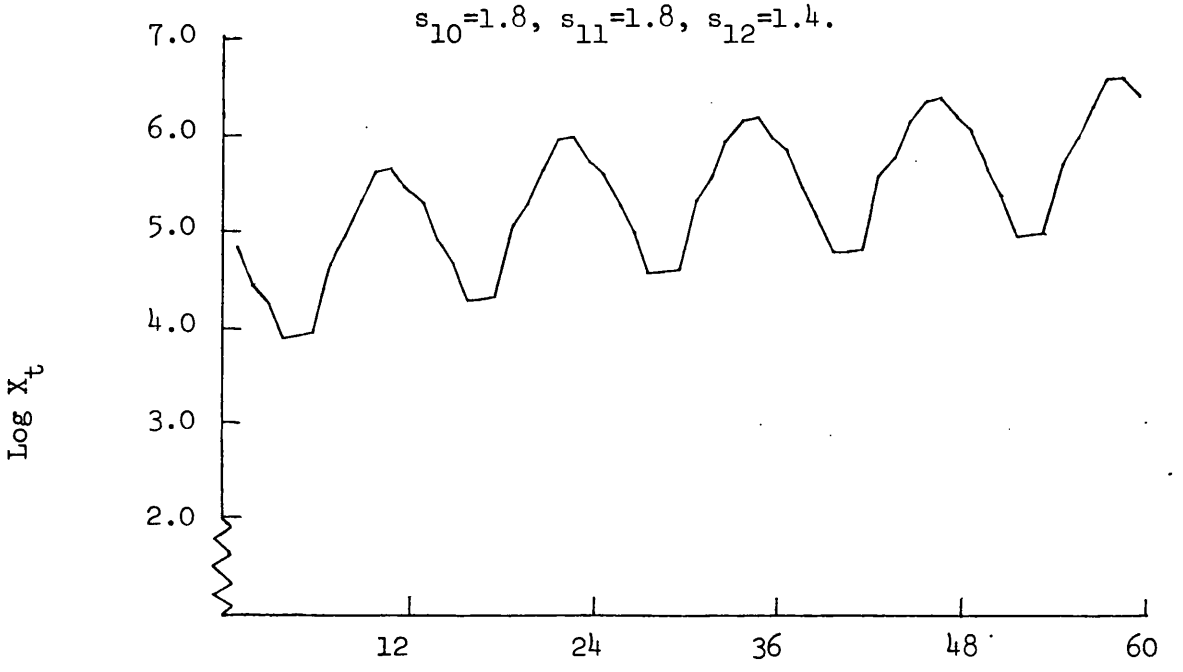


FIGURE 4.1(b)

The logarithmic transformation of the above series

in Figure 4.1(b). This latter series appears to exhibit a non-linear trend over the first twelve to eighteen months data but thereafter the trend seems to be fairly linear. This behaviour can be explained by referring to the logarithmic expansion (4.2.5). Initially, the ratio $\beta/\alpha(t)$ takes the value $\frac{5}{100}$ $(= \frac{1}{20})$ so that the expansion (4.2.5) is valid only for $|u|<20$ and the linear approximation will hold good only for u = 1, 2 or 3 at the most, i.e. the initial trend is not even locally linear. At a later point in time, however, the ratio $\beta/\alpha(t)$ will be much smaller. For example, when t = 40, $\alpha(40) = 300$ and so $\beta/\alpha(t) = \frac{5}{300}$ $(= \frac{1}{60})$. Thus the logarithmic function can be expanded for $|u|<60$ and the trend will be approximately linear over a wider range of values of u.

Using the same values for α and s_j (j = 1, 2, 3,...,12) but with $\beta = 1$, a second series was generated and a logarithmic transformation was again applied. These two series are shown in Figures 4.2(a) and (b). On this occasion, $\beta/\alpha(o) = \frac{1}{100}$, the logarithmic expansion can be used for $|u|<100$ and the linear approximation will be quite reasonable over a fair range of values for u. Thus, as can be seen from Figure 4.2(b), the trend will be locally linear over the earlier part of the series as well as in the latter part.

We have seen that when a logarithmic transformation is applied to a series described by a model of the form (4.2.3) the resulting series may possess a good approximation to a linear trend. This leads naturally to the question: What form must the trend take in the original model in order for a logarithmic transformation to produce a trend which is exactly linear? By assuming that the original model is

$$X_{t+u} = T_u s_{t+u} \qquad 4.2.6$$

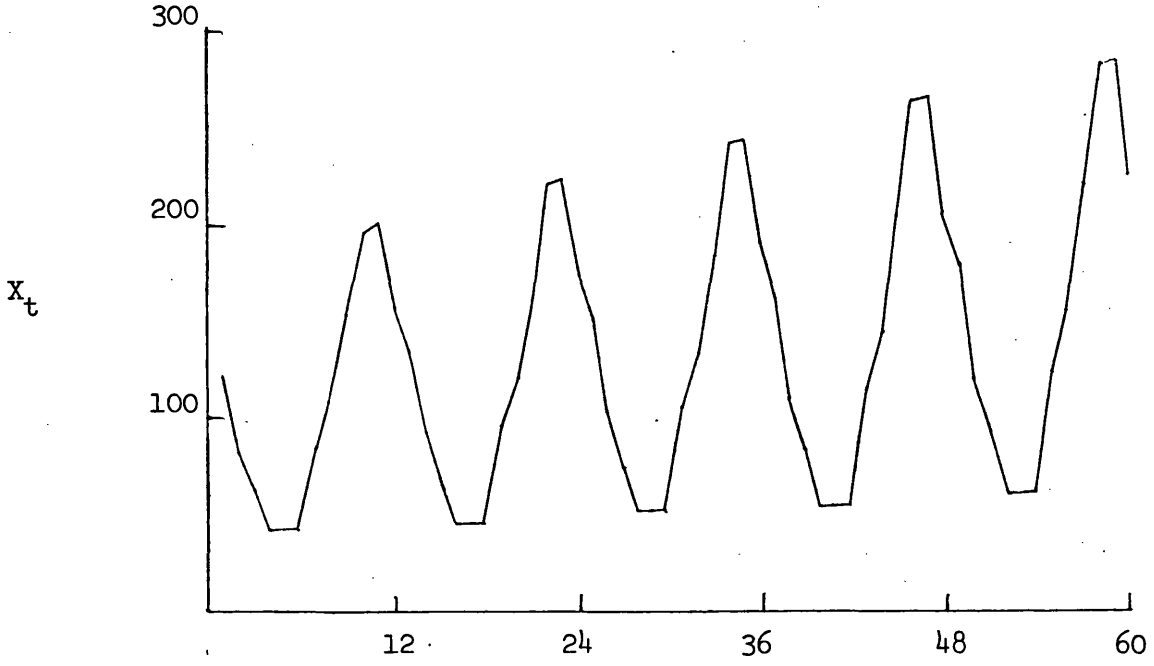


FIGURE 4.2(a)

A purely deterministic series given by

$X_t = (\alpha+\beta t)s_t$, where $\alpha=100$, $\beta=1$,
$s_1=1.2$, $s_2=0.8$, $s_3=0.6$, $s_4=0.4$,
$s_5=0.4$, $s_6=0.4$, $s_7=0.8$, $s_8=1.0$,
$s_9=1.4$, $s_{10}=1.8$, $s_{11}=1.8$, $s_{12}=1.4$.

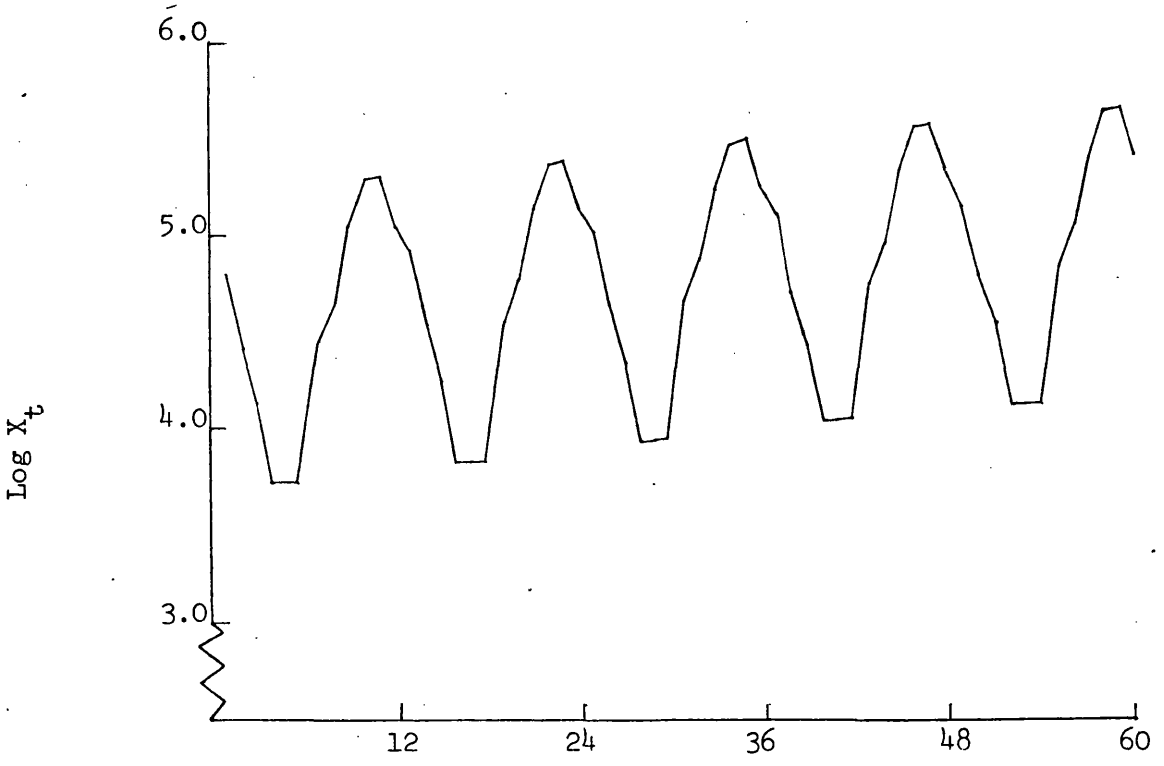


FIGURE 4.2(b)

The logarithmic transformation of the above series

where T_u represents the trend component at time t, it is easy to show that T_u must take the form

$$T_u = Ce^{Du} \qquad 4.2.7$$

where C and D are constants. This point was also made by Professor P.J. Harrison during the discussion on the paper by Chatfield and Prothero (1973a).

In conclusion, the effect of a logarithmic transformation on a series described by a linear trend and multiplicative seasonal variation depends on the monthly growth rate at any given time. The study of a series similar to the Company X data revealed that while for the most part the transformed series possessed a local linear trend the same was not true for the early part of the series. In the next section we shall see if this absence of local linearity in the first 12 - 18 months data was responsible for the poor forecasts obtained in Chapter 3.

4.3 Analysis of the Company X Data Using a Different Transformation

As mentioned in Section 4.1, Wilson (1973) analysed the Company X data using a transformation of the type proposed by Box and Cox (1964). Such a transformation is designed to produce linearity in the transformed series. This alternative analysis will not be described in detail but some of the results derived from it will be compared with those generated in Chapter 3 using model (A).

Wilson (1973) assumed a model of a similar form to model (A) and the parameters, including the transformation parameter, were estimated from the first 60 observations. The resulting fitted model was

$$(1 + 0.37B)\nabla\nabla_{12}X_t^{0.34} = (1 - 0.79B^{12})a_t \qquad 4.3.1$$

Re-estimating the parameters, using all of the data, by the approach to be described in section 4.4 we found that the fitted model was

$$(1 + 0.50B)\ \nabla\nabla_{12}X_t^{0.23} = (1 - 0.80B^{12})a_t \qquad 4.3.2$$

The fitted model (4.3.2) is not quoted by Wilson (1973).

Setting $z_t = X_t^{0.23}$ in equation (4.3.2), the forecasts $\hat{z}_t(\ell)$ were computed from May 1971 for $\ell = 1, 2, 3, \ldots, 12$. The point forecasts $\hat{X}_t(\ell)$ from May 1971 were obtained from

$$\hat{X}_t(\ell) = (\hat{z}_t(\ell))^{\frac{1}{0.23}} \qquad 4.3.3$$

since the bias (see Granger and Newbold (1970)) involved in using such a forecast was found to be less than 5% for all lead times. The latter forecasts are plotted in Figure 4.3 and tabulated together with tolerance limits in Table 4.1. Also shown in Figure 4.3 and Table 4.1 are the corresponding forecasts resulting from the fitted model (A) (equation (3.5.1)) based on the logarithmic transformation employed in Chapter 3.

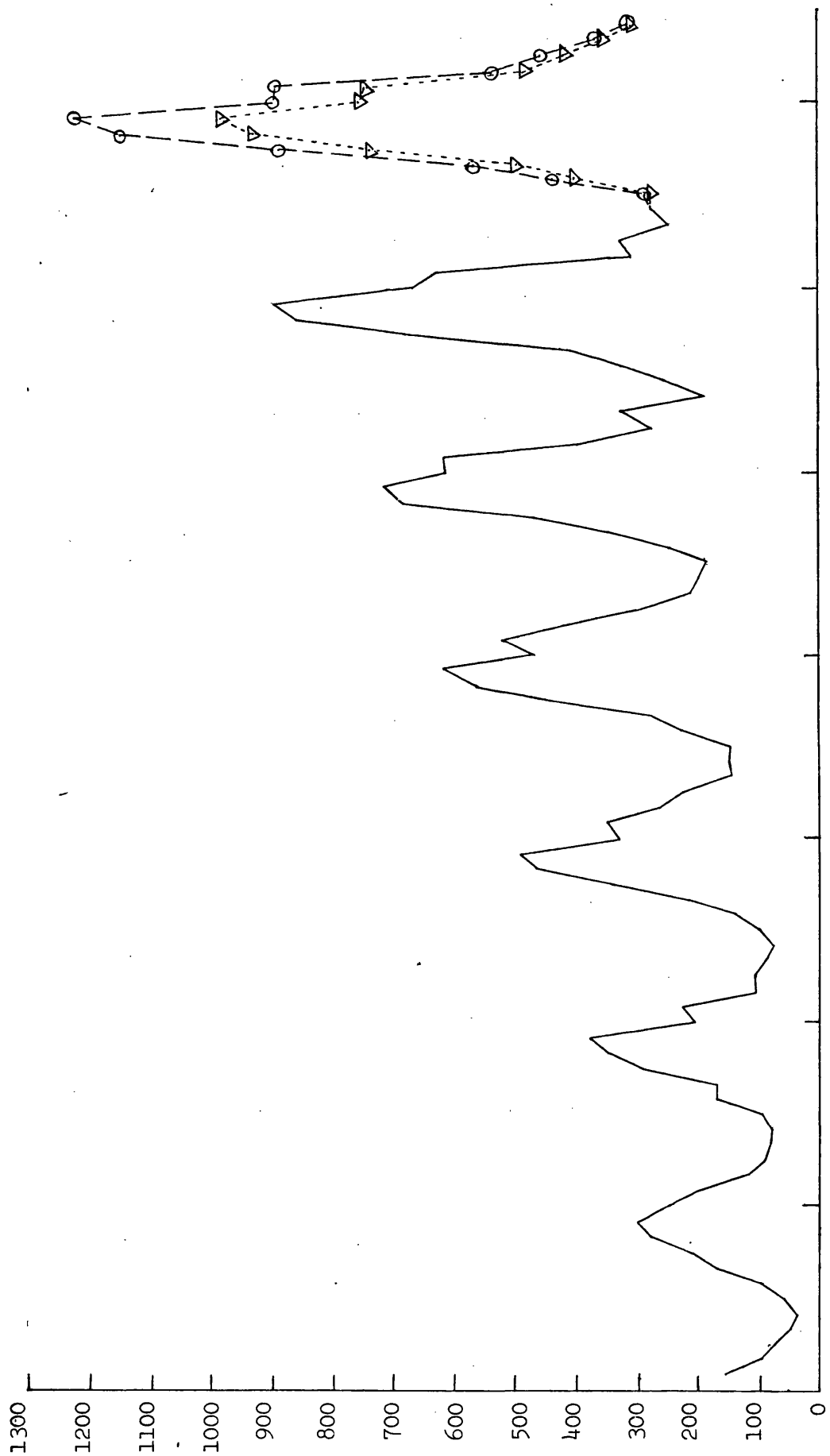


FIGURE 4.3

Forecasts made at May 1971 using model $(1+0.47B)\nabla\nabla_{12}\log_{10} X_t = (1-0.81B^{12})a_t$ (o--o--o) and model $(1+0.50B)\nabla\nabla_{12} X_t^{0.23} = (1-0.80B^{12})a_t$ (▲----▲----▲)

Table 4.1 Forecasts Made At May 1971 For Lead Times 1 - 12

Lead Time	Model (A)	Model (4.3.2)
1	286 ± 103	275 ± 83
2	437 ± 186	399 ± 125
3	562 ± 315	493 ± 176
4	881 ± 521	734 ± 263
5	1148 ± 759	929 ± 350
6	1221 ± 899	980 ± 393
7	897 ± 721	751 ± 344
8	889 ± 735	742 ± 363
9	535 ± 471	482 ± 276
10	452 ± 412	416 ± 260
11	367 ± 336	352 ± 240
12	314 ± 300	310 ± 228

It can be seen that overall the point forecasts generated by the model (4.3.2) differed considerably from those resulting from the model (A) fitted to the logarithmic transformation of the original data. This difference is most noted in the peaks of the seasonal cycle and rather less marked in the troughs.

Thus it seems that the departure from linearity in the early part of the logarithmic transformation of the Company X data was responsible for the poor forecasts produced by model (A). Wilson (1973) showed that by expressing model (A) in terms of past observations only, a considerable amount of weight was given to observations in the first years data. On the other hand, the point forecasts computed using the model (B) introduced in Section 3.7 depended only on the most recent 26 observations where the assumption of linearity on the logarithmic scale was quite acceptable. This

explains why model (B) produced more reasonable point forecasts than model (A). However, the tolerance limits associated with model (B) were of a similar magnitude to those of model (A). From Table 4.1 it can be seen that model (4.3.2) gives rise to much narrower tolerance limits than those resulting from model (A). Hence the absence of linearity on the logarithmic scale at the start of the series appears to be the cause of the wide tolerance limits associated with the models (A) and (B) and the unreasonable point forecasts produced by the former.

The above comparison is based on just one set of forecasts from one particular origin. To achieve a more general comparison, the fitted model (4.3.1) and the model (A), fitted to the first 60 observations, were used to generate forecasts over the remaining 17 observations. The fact that the models were based on different transformations made a comparison rather difficult. However some measure of the relative forecasting potential of the two models was achieved by calculating the mean absolute forecast errors in the original variable X_t. These quantities are given in Table 4.2 for lead times 1 and 6.

Table 4.2 Mean Absolute Forecast Errors

Lead Time	Model (A)	Model (4.3.1)
1	51.8	49.2
6	124.3	79.6

For both lead times quoted in Table 4.2 the model (4.3.1) produced the smaller mean absolute forecast errors. The difference between the forecasting performance of the two models was small in the case of the lead time 1 forecasts but for lead time 6, model

(4.3.1) reduced the mean absolute error by about a third. From this it can be concluded that the choice of transformation is not too important with respect to the one step ahead forecasts but for higher lead times the consequences of any lack of linearity in the transformed series become progressively more serious.

4.4 The Use of Non-linear Transformations in Time Series Analysis

The transformation employed by Wilson (1973) in the model (4.3.1) is taken from the set of non-linear transformations discussed by Box and Cox (1964). The general non-linear transformation takes the form

$$z_t = \begin{cases} (X_t + m)^{\lambda} & \lambda \neq 0 \\ \log (X_t + m) & \lambda = 0 \end{cases} \qquad 4.4.1$$

where the parameter m is chosen so that $X_t + m$ is positive for all t. For simplicity, it will be assumed that X_t is positive for all t so that equation (4.4.1) can be replaced by

$$z_t = \begin{cases} X_t^{\lambda} & \lambda \neq 0 \\ \log X_t & \lambda = 0 \end{cases} \qquad 4.4.2$$

Now in choosing a suitable value for the transformation parameter λ there are several approaches which could be adopted.

Judging by some of the published Box-Jenkins analyses, e.g. Box and Jenkins (1970), Makridakis and Wheelwright (1972), Tomasek (1972), it would appear that a subjective choice of transformation can often be made (usually $\lambda = 0$ or $\lambda = 1$). Alternatively, λ can be estimated from the data using the techniques described in Box and Cox (1964). The essential features of this approach are now outlined.

The vector $\underline{X}' = (X_{-N+n+1}, X_{-N+n+2}, X_{-N+n+3},\ldots,X_n)$ is used to denote the N observations which compose the non-stationary, seasonal time series X_t. A transformation of the form defined by equation (4.4.2) is to be applied to the series X_t, the observations of the transformed series z_t constituting the vector $\underline{z}' = (z_{-N+n+1}, z_{-N+n+2}, z_{-N+n+3},\ldots,z_n)$. It is assumed that z_t can be described by the general multiplicative seasonal model (2.6.5), or

$$\phi_p(B)\ \Phi_p(B^s)w_t = \theta_q(B)\ \Theta_Q(B^s)a_t \qquad 4.4.3$$

where w_t $(t = 1,2,3,\ldots,n)$ represents the stationary series resulting from differencing z_t the appropriate number of times. The vector $w_n' = (w_1, w_2, w_3,\ldots w_n)$ denotes the observations of w_t while the $(p + P)$ x 1 vector $\underline{\phi}$ and the $(q + Q)$ x 1 vector $\underline{\theta}$ refer to the respective sets of autoregressive and moving average parameters.

Following the approach of Box and Cox (1964) the likelihood associated with the X_t's for a fixed value of λ is

$$L'(\underline{\phi},\underline{\theta},\sigma_a|\underline{X}) = L(\underline{\phi},\underline{\theta},\sigma_a|\underline{z})|J| \qquad 4.4.4$$

where $L(\underline{\phi},\underline{\theta},\sigma_a|\underline{z})$ is the likelihood associated with the zeries z_t and J is the Jacobian of the transformation from the z_t's to the X_t's. For $\lambda \neq 0$,

$$J = \begin{vmatrix} \lambda X_{-N+n+1}^{\lambda-1} & & & & \\ & \lambda X_{-N+n+2}^{\lambda-1} & & & \\ & & \lambda X_{-N+n+3}^{\lambda-1} & & \\ & & & \ddots & \\ & & & & \lambda X_{n}^{\lambda-1} \end{vmatrix}$$

$$= \lambda^{N} \prod_{t=-N+n+1}^{n} X_t^{\lambda-1} \qquad 4.4.5$$

so that equation (4.4.4) becomes

$$L'(\underline{\phi},\underline{\theta},\sigma_a|\underline{X}) = L(\underline{\phi},\underline{\theta},\sigma_a|\underline{z})|\lambda|^{N} \prod_{t=-N+n+1}^{n} X_t^{\lambda-1} \qquad 4.4.6$$

and the log-likelihood is

$$\ell'(\underline{\phi},\underline{\theta},\sigma_a|\underline{X}) = \ell(\underline{\phi},\underline{\theta},\sigma_a|\underline{z}) + N \log|\lambda| + (\lambda-1) \sum_{t=-N+n+1}^{n} \log X_t \qquad 4.4.7$$

where $\ell(\underline{\phi},\underline{\theta},\sigma_a|\underline{z}) = \log[L(\underline{\phi},\underline{\theta},\sigma_a|\underline{z})]$.

Now using the result quoted by Box and Jenkins (1970, page 273),

$$L(\underline{\phi},\underline{\theta},\sigma_a|\underline{z}) = (2\pi\sigma_a^2)^{-\frac{n}{2}} G(\underline{\phi},\underline{\theta}) e^{-\frac{1}{2\sigma_a^2} S_\lambda(\underline{\phi},\underline{\theta})} \qquad 4.4.8$$

and

$$\ell(\underline{\phi},\underline{\theta},\sigma_a|\underline{z}) \propto -\frac{n}{2}\log\sigma_a^2 + \log G(\underline{\phi},\underline{\theta}) - \frac{1}{2\sigma_a^2} S_\lambda(\underline{\phi},\underline{\theta}) \qquad 4.4.9$$

where $G(\underline{\phi},\underline{\theta})$ is some function of the parameters $\underline{\phi}$ and $\underline{\theta}$ and

$$S_\lambda(\underline{\phi},\underline{\theta}) = \sum_{t=-\infty}^{n} [a_t|\underline{z},\underline{\phi},\underline{\theta}]^2 \qquad 4.4.10$$

For moderate or large values of n the term log $G(\underline{\phi},\underline{\theta})$ is unimportant compared with the other quantities in the expression (4.4.9) and so for most practical purposes substitution for $\ell(\underline{\phi},\underline{\theta},\sigma_a|\underline{z})$ in equation (4.4.7) results in

$$\ell'(\underline{\phi},\underline{\theta},\sigma_a|\underline{X}) \propto -\frac{n}{2}\log \sigma_a^2 - \frac{1}{2\sigma_a^2} S_\lambda(\underline{\phi},\underline{\theta}) + N \log |\lambda| + (\lambda-1) \sum_{t=-N+n+1}^{n} \log X_t \qquad 4.4.11$$

Differentiating $\ell'(\underline{\phi},\underline{\theta},\sigma_a|\underline{X})$ partially with respect to σ_a^2 and equating to zero leads to the maximum likelihood estimate for σ_a^2

$$\hat{\sigma}_a^2 = \frac{S_\lambda(\hat{\underline{\phi}},\hat{\underline{\theta}})}{n} \qquad 4.4.12$$

(see Box and Jenkins (1970, page 277)) where $S_\lambda(\hat{\underline{\phi}},\hat{\underline{\theta}})$ is the minimum unconditional sum of squares for a fixed value of λ. Thus the value of the log-likelihood maximised with respect to $\underline{\phi}$, θ and σ_a^2 is, apart from a constant, given approximately by

$$\ell'_{max}(\underline{\phi},\underline{\theta},\sigma_a|\underline{X}) \simeq -\frac{n}{2}\log \frac{S_\lambda(\hat{\underline{\phi}},\hat{\underline{\theta}})}{n} + N \log |\lambda| + (\lambda-1) \sum_{t=-N+n+1}^{n} \log X_t \qquad 4.4.13$$

However, Dr. Wilson has pointed out that the form (4.4.13) is affected by scaling. If all the terms of X_t are multiplied by some fixed constant K, then all the terms of z_t are multiplied by K^λ. Consequently the first term of $\ell'_{max}(\underline{\phi},\underline{\theta},\sigma_a|\underline{X})$ is affected by $-n\lambda \log K$ and the last term by $N(\lambda-1) \log K$ leaving a net effect $(-n\lambda + N\lambda - N) \log K$ and so the maximised log-likelihood (4.4.13) is affected by scaling. The reason for this is that the likelihood (4.4.8) is, strictly speaking, associated with the series of n values $z_1, z_2, z_3, \ldots, z_n$

rather than with the series of N values z_{-N+n+1}, z_{-N+n+2}, $z_{-N+n+3},\ldots,z_n$. Thus the likelihood corresponding to the series z_{-N+n+1}, z_{-N+n+2}, $z_{-N+n+3},\ldots,z_n$ should be

$$L(\underline{\phi},\underline{\theta},\sigma_a|\underline{z}) = (2\pi\ \sigma_a^2)^{-\frac{N}{2}}\ G(\underline{\phi},\underline{\theta})e^{-\frac{1}{2\sigma_a^2}\ S_\lambda(\underline{\phi},\underline{\theta})}$$

although since the unconditional sum of squares $S_\lambda(\underline{\phi},\underline{\theta})$ is computed via the n values $w_1,w_2,w_3,\ldots,w_n$ the estimate of σ_a^2 is still given by equation (4.4.12). Hence the first term in the expression (4.4.13) should include the factor $\frac{N}{2}$ and not $\frac{n}{2}$, i.e.

$$\ell'_{max}(\underline{\phi},\underline{\theta},\sigma_a|\underline{X}) \simeq -\frac{N}{2}\log\ \frac{S_\lambda(\hat{\underline{\phi}},\hat{\underline{\theta}})}{n}$$

$$+ N \log|\lambda| + (\lambda-1) \sum_{t=-N+n+1}^{n} \log X_t \qquad 4.4.14$$

The form (4.4.14) is now unaffected by scaling.

Similarly when $\lambda = 0$, $J = \prod_{t=-N+n+1}^{n} \frac{1}{X_t}$ and

$$\ell'_{max}(\underline{\phi},\underline{\theta},\sigma_a|\underline{X}) \simeq -\frac{N}{2}\log \frac{S_o(\hat{\underline{\phi}},\hat{\underline{\theta}})}{n}$$

$$- \sum_{t=-N+n+1}^{n} \log X_t \qquad 4.4.15$$

By plotting $\ell'_{max}(\underline{\phi},\underline{\theta},\sigma_a|\underline{X})$ against λ over a range of values for λ, an estimate of the transformation parameter can be obtained. The estimate corresponds to the value of λ for which the function $\ell'_{max}(\underline{\phi},\underline{\theta},\sigma_a|\underline{X})$ is maximised.

An example of the above procedure is given by Wilson (1973) who obtained a value of $\hat{\lambda} = 0.34$ using the first 60 observations of the Company X data and assuming a model of order $(1,1,0) \times (0,1,1)_{12}$.

A closer examination of the use of the methods of Box and Cox

(1964) in time series analyses in general and in conjunction with the Box-Jenkins forecasting procedure in particular yielded a number of interesting points. These will now be discussed.

In estimating the transformation parameter λ, we have seen that the form of the model must be identified before the log-likelihood function can be computed. However, in identifying the order of an A.R.I.M.A. model, use is made of the sample auto-correlation function. It is not clear whether the sample auto-correlation function of the untransformed data or of some other specific transformation of the data (e.g. the logarithmic transformation in the case of the Company X data) should be examined. That the sample autocorrelation function is not invariant under transformation can be seen from Table 4.3 where estimated auto-correlations of the differenced Company X data $w_t(=\nabla\nabla_{12} X_t^{\lambda})$ are shown for lags 1 - 12 and for values of λ between 0 and 1. This point has been noted by Chatfield and Prothero (1973a) in replying to the discussion on the original paper.

In Chapter 3 we saw that when $\lambda = 0$, an inspection of the sample autocorrelation of w_t suggested the model (3.3.5) (model (A)). For $\lambda = 1$, the sample autocorrelation function (see Table 4.3 for lags 1 - 12) leads to the tentative model

$$(1-\phi B)w_t = (1-\theta B)(1-\theta B^{12})a_t \qquad 4.4.16$$

Thus, for a fixed degree of differencing, one strategy would be to calculate the sample autocorrelation function over a range of values for λ and then to assume the most general identified model when estimating λ. This solution would of course not apply to situations where the identified model for one value of λ is based on a different degree of differencing to the model implied for another value of λ. It is however possible that the estimate of the transformation parameter

Table 4.3 Sample autocorrelation functions for differenced Company X data $\nabla\nabla_{12}X_t$, λ = 0 to 1, lags 1-12.

λ \ Lag	1	2	3	4	5	6	7	8	9	10	11	12
0	-0.58	0.36	-0.22	0.05	-0.05	0.10	-0.17	-0.02	0.10	-0.26	0.44	-0.36
0.1	-0.59	0.36	-0.22	0.06	-0.06	0.11	-0.17	-0.02	0.11	-0.26	0.44	-0.36
0.2	-0.60	0.36	-0.22	0.06	-0.07	0.11	-0.17	-0.02	0.11	-0.26	0.44	-0.35
0.3	-0.59	0.36	-0.22	0.06	-0.09	0.11	-0.17	-0.02	0.12	-0.26	0.45	-0.34
0.4	-0.59	0.35	-0.22	0.06	-0.10	0.11	-0.17	-0.01	0.12	-0.25	0.45	-0.33
0.5	-0.57	0.35	-0.22	0.06	-0.11	0.11	-0.17	-0.01	0.12	-0.24	0.45	-0.32
0.6	-0.56	0.34	-0.22	0.05	-0.13	0.11	-0.18	0.00	0.12	-0.23	0.45	-0.31
0.7	-0.53	0.33	-0.22	0.05	-0.14	0.11	-0.19	0.00	0.12	-0.22	0.45	-0.29
0.8	-0.51	0.32	-0.23	0.04	-0.15	0.11	-0.19	0.00	0.12	-0.20	0.45	-0.27
0.9	-0.48	0.31	-0.24	0.02	-0.17	0.11	-0.20	0.00	0.12	-0.18	0.45	-0.25
1.0	-0.44	0.30	-0.24	0.01	-0.18	0.10	-0.21	0.00	0.11	-0.17	0.45	-0.23

will not be greatly affected by the choice of model. To look at this possibility further, the function $\ell'_{max}(\underline{\phi},\underline{\theta},\sigma_a|\underline{X})$ was calculated for three different A.R.I.M.A. models. The three models entertained were

$$(1-\phi B)\ \nabla\nabla_{12}\ X_t^{\lambda} = (1-\Theta B^{12})a_t \qquad 4.4.17$$

$$(1-\phi B)(1-\Phi B^{12})\ \nabla\nabla_{12}X_t^{\lambda} = a_t \qquad 4.4.18$$

and

$$(1-\phi_1 B - \phi_2 B^2)\ \nabla_{12}^2\ X_t^{\lambda} = (1-\Theta_1 B^{12} - \Theta_2 B^{24})a_t \qquad 4.4.19$$

Models (4.4.17) and (4.4.18) possess the same forms as models (A) and (B) introduced in Section 3.7. Model (4.4.17) was chosen because it was the initially identified model when λ = 0 (see Chapter 3) and model (4.4.18) because it produced quite reasonable point forecasts from May 1971 when λ = 0 (again see Chapter 3). The model

(4.4.19) was selected because it is based on a different differencing operator to the other two. A model of the form (4.4.19) was suggested by Wilson (1973) as being suitable for describing the untransformed data, i.e. $\lambda = 1$.

The log-likelihood functions for each of the three models, computed from the whole of the data, are plotted in Figure 4.4. Also indicated are the point estimates for λ and 95% confidence intervals for λ, obtained using the method set out in Box and Cox (1964).

The most striking feature of Figure 4.4 is that the point estimates for λ are very close to one another (all lying between 0.2 and 0.3) suggesting that the transformation is not influenced to any great extent by the choice of model. Although the point estimates are in close agreement, the confidence intervals for these three estimates are more variable. The log-likelihood function for the model (4.4.17) possesses a much sharper maximum than the log-likelihoods associated with the other two models, indicating a more precise point estimate. In fact, the 95% confidence interval for λ in the case of model (4.4.17) ranges from 0.09 to 0.36 while for model (4.4.18) it is from 0.07 to 0.46 and for model (4.4.19) from 0.01 to 0.52. It can also be seen from Figure 4.4 that a value of $\lambda = 0$, i.e. a logarithmic transformation, is rejected with rather more confidence for model (4.4.17) than for model (4.4.18) which explains why, in Chapter 3, model (B) proved more acceptable than model (A). Finally, for model (4.4.19) a logarithmic transformation is only just rejected at the 5% level of significance. This is rather surprising in view of the fact that Wilson (1973) fitted a model of this form to the untransformed data. More comments on the use of this model will be made in Section 4.5.

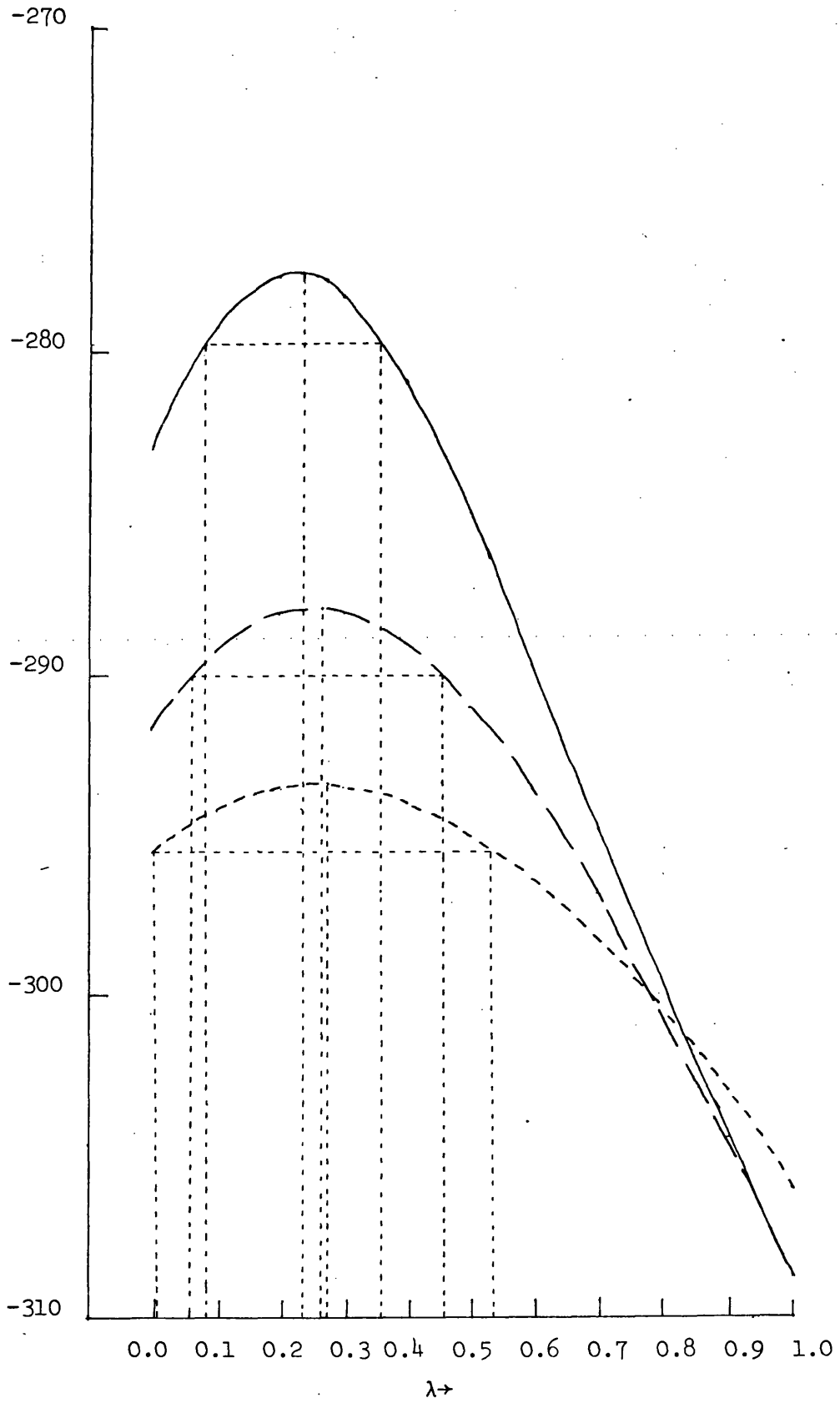


FIGURE 4.4

Log likelihood functions for various models.

$(1-\phi B)\nabla\nabla_{12} X_t^{\lambda} = (1-\theta B^{12})a_t$ ————————

$(1-\phi B)(1-\Phi B^{12})\nabla\nabla_{12} X_t^{\lambda} = a_t$ —— —— —— ——

$(1-\phi_1 B-\phi_2 B^2)\nabla_{12}^2 X_t^{\lambda} = (1-\theta_1 B^{12}-\theta_2 B^{24})a_t$ - - - - - - - -

We now look briefly at how a relatively small number of observations can influence the estimation of a transformation parameter in the field of time series analysis. Again for illustrative purposes reference will be made to the data of Company X.

Fitting the model (4.4.17) to the first 5 years data (60 observations), Wilson (1973) estimated λ to be 0.34 and a logarithmic transformation was rejected at the 5% level. As can be seen from Figure 4.4, on fitting the same model to all 77 observations it was found that $\hat{\lambda} = 0.23$ and although $\lambda = 0$ was again rejected, a value of $\lambda = 0.1$ was just inside the 95% confidence interval. Thus it may well be that as more observations become available a logarithmic transformation will prove acceptable - a conclusion arrived at in Section 4.2. Further confirmation of the theory developed in Section 4.2 can be obtained by fitting the model (4.4.17) to the final 65 observations of the Company X data, i.e. by excluding the first years data. An estimate of λ was found to be 0.16 and the 95% confidence interval for λ included the logarithmic transformation.

Just how critical the estimation of a transformation parameter can be, is emphasised by the fact that the value of λ estimated from the first 60 observations (assuming model (4.4.17)) is only just within the 95% confidence interval for λ obtained using all the data.

4.5 General Summary and Conclusions

In view of some of the results obtained in this chapter it is worth re-stating the reasons why a logarithmic transformation was applied to the Company X data in Chapter 3. An inspection of the original data revealed that the series possessed an approximate linear trend and a multiplicative seasonal variation so that a non-linear transformation was called for. Wishing, if possible,

to avoid the added complications involved in estimating a transformation by the methods of Box and Cox (1964), it was tentatively decided to employ a logarithmic transformation. Further, reference to Box and Jenkins (1970, page 94) suggested that for "the sales of a recently introduced commodity" where the "sales volume was increasing at a rapid rate and that it was the percentage fluctuations which showed non-stationary stability" "it would clearly be sensible to analyse the logarithm of sales".

A graph of the transformed series (Figure 3.2) showed that the latter exhibited an approximate linear trend and additive seasonal component, although the trough in the first years data was rather low. The logarithms of the monthly sales were plotted individually for each month (Figure 3.3) and the resulting lines were found to be roughly linear and parallel. Thus at this stage there seemed to be no obvious reason why any other transformation should be contemplated.

However, the forecasts generated by the model (A) based on the logarithmic transformation compared unfavourably with those generated by Wilson (1973) from the same model using a different transformation. This led naturally to the conclusion that the logarithmic transformation was responsible for the disappointing results obtained in Chapter 3. In order to find out why this should be so, a logarithmic transformation was applied to a series exhibiting a deterministic linear trend and multiplicative seasonal variation, i.e. a series similar to the Company X data. The conclusion arrived at was that the transformed series will generally possess a local non-linear trend. Nevertheless, provided the monthly growth rate is small, the assumption of a local linear trend will be valid to a good approximation. In the case of the Company X data, only in the first 12 - 18 months was the growth rate too high to justify this assumption.

The effect of the initial absence of linearity in the transformed series on the derived forecasts has been illustrated in Section 4.3 where the forecasting performance of the same model based on a logarithmic transformation and on a transformation estimated from the data by the methods of Box and Cox (1964) was assessed. Generally speaking, the one step ahead forecasts were not greatly affected by the transformation employed but for higher lead times the choice of transformation became more critical.

Although the Box-Cox transformations are specifically designed to produce linearity in the transformed series, their use does present certain difficulties. The estimation of the transformation parameter is a rather long, complicated procedure which necessitates the identication of an A.R.I.M.A. model before any estimate can be computed. This latter point in itself creates a problem since the main identification tool, the sample autocorrelation function, is not invariant under transformations. In the case of the Company X data it was however found that the estimate of the transformation parameter was not greatly altered by assuming different models.

A cruder method of estimating the transformation parameter has been used by Box and Jenkins (1973). This involves an inefficient trend estimate (as pointed out by Chatfield and Prothero (1973b)) and so the parameter estimate may tend to be rather unreliable.

Apart from the problems involved in estimating λ, two other points require comment. Firstly, the use of a transformation of the form $z_t = X_t^{\lambda}$ makes interpretation of the fitted A.R.I.M.A. model in terms of the original data even more difficult. Secondly, for short or medium length series the estimate of λ can change considerably over short periods of time so that the need for re-estimation should always be considered.

In the light of the above problems and criticisms it is perhaps pertinent to seek some alternative to employing a non-linear trans-

formation. One possibility is an approach similar to that adopted by Wilson (1973) in proposing the use of the seasonal A.R.I.M.A. model of order $(2,0,0) \times (0,2,2)_{12}$ for describing the Company X data. Although Wilson (1973) fitted this model to the untransformed data, the Box-Cox estimate of the transformation parameter was $\hat{\lambda} = 0.26$. Nevertheless, the log-likelihood function associated with this model is flatter than those for the other two models shown in Figure 4.4 and in general it would seem to be advisable to avoid the complications involved in employing a non-linear transformation by applying a suitable choice of differencing operator to the untransformed observations, wherever such a choice is possible.

CHAPTER 5

DIFFERENCING

5.1 Introduction

An important step at the identification stage of the Box-Jenkins forecasting procedure involves the selection of the degree of differencing which will reduce some suitably transformed non-stationary series z_t to a stationary series w_t. Indeed, according to Akaike (1973), "when the variation of the systematic part, i.e. the trend and seasonality, is dominant the effectiveness of the A.R.I.M.A. model is mainly determined by the initial simple differencing operations and not by the time-consuming A.R.M.A. model fitting to the stationary part".

In this chapter we will discuss the choice of the differencing operator and then examine some of the consequences of using differencing as a means of achieving stationarity. We will concentrate on seasonal time series and for convenience the seasonal cycle is assumed to have a period 12.

Section 5.2 will show how classical (or traditional) time series models can often be utilised to suggest the appropriate degree of differencing. This approach could be considered to be complementary to the usual methods advocated by Box and Jenkins (1970, Chapter 6). The latter will also be outlined in Section 5.2.

The desirability of keeping track of the variance of the differenced series for successive degrees of differencing will be proposed in Section 5.3 while the effect of the differencing operation on the error components in the classical model will be considered in Section 5.4.

In section 5.5 we will see how in some cases the choice of differencing operator can influence the tolerance limits attached to a set of point forecasts.

The conclusions drawn from Sections 5.2 to 5.5 will be summarised

in Section 5.6.

5.2 Selecting the Degree of Differencing

Let us begin this section by describing the approach to selecting the degree of differencing necessary to produce stationarity, recommended by Box and Jenkins (1970, Chapter 6) and outlined in Section 2.4.1.

5.2.1 The Box-Jenkins Approach

Box and Jenkins (1970, page 75) show that for a non-seasonal, stationary mixed autoregressive moving average process of order (p,q), the theoretical autocorrelation function $\rho(k)$ satisfies the difference equation

$$\rho(k) - \phi_1 \rho(k-1) - \phi_2 \rho(k-2) - .... - \phi_p \rho(k-p) = 0$$

for $k \geq q + 1$, or

$$\phi_p(B)\, \rho(k) = 0 \qquad 5.2.1$$

Letting $\phi_p(B) = \prod_{i=1}^{p} (1 - G_i B)$, then providing the roots G_i are distinct, equation (5.2.1) has a solution of the form

$$\rho(k) = A_1 G_1^k + A_2 G_2^k + A_3 G_3^k + ... + A_p G_p^k \qquad 5.2.2$$

for $k > q - p$,

where the A_i's ($i = 1,2,3,\ldots,p$) are constants.

In order for the stationarity condition defined by Box and Jenkins (1973, page 74) to be satisfied, the roots G_i ($i = 1,2,3,\ldots,p$) must lie inside the unit circle. Thus, in the case of a stationary model when none of the roots ~~are~~ is close to the boundary of the unit circle, the function $\rho(k)$ will die out quickly for moderate and large values of k. However, if one of the roots, e.g. G_1, approaches unity then Box and

Jenkins (1970, page 74) show that $\rho(k)$ will not die out quickly. Failure of the autocorrleation function to die out quickly would therefore tend to indicate the presence of a root close to unity, i.e. non-stationarity.

In practice of course the theoretical autocorrelation function would not be known and so the behaviour of the sample autocorrelation function would be examined. The Box-Jenkins approach is therefore to inspect the sample autocorrelation function for successive differences of the non-stationary series, until stationarity is indicated by the behaviour mentioned in the previous paragraph.

For a seasonal process z_t, the procedure is similar to that described above except that the sample autocorrelation function of the series $\nabla^d \nabla_{12}^D$ is examined for values of d = 0,1,2,..., D = 0,1,2,... . Non-stationarity with respect to the seasonal period will be characterised by the failure of the autocorrelations at lags 12,24,36,..., to die out quickly.

5.2.2 The Use of Classical Time Series Models

The degree of differencing necessary to produce stationarity can often be decided by assuming that the series in question can be decomposed into three components: a trend (possibly local), a seasonal component and an error term. The three usual forms of this so called classical representation are given by Kendall (1973, page 56) but we shall confine ourselves to the two cases in which the error (or unpredictable) component is additive with respect to the other two components. Algebraically, it will be assumed that a series X_t can be described by the model

$$X_t = d_t + n_t \qquad 5.2.3$$

where d_t is a deterministic component composed of trend and seasonality

and n_t is the error component (not generally independent)..

If the variation associated with the deterministic part is dominant then the differencing operator capable of reducing X_t to a stationary series would be one which removes the deterministic components from X_t i.e. one which renders

$$\left.\begin{aligned} &\nabla^d \nabla_{12}^D d_t = 0 \\ \text{or} \quad &\\ &\nabla^d \nabla_{12}^D d_t = C \end{aligned}\right\} \qquad 5.2.4$$

where C is a constant.

In practice, the approximate form of d_t must be deduced by inspecting the series under consideration. Firstly, one must decide whether the seasonal component is additive (the amplitude of the seasonal cycle is independent of the level of the series) or multiplicative (the amplitude of the seasonal cycle increases proportionally with the level of the series). Secondly, an approximate trend must be specified.

We shall now look more closely at two particular forms of d_t, viz

$$d_t = \alpha + \beta t + s_t \qquad 5.2.5$$

and

$$d_t = (\alpha + \beta t)s_t \qquad 5.2.6$$

where α, β and s_j ($j = 1,2,3,\ldots,12$) are constants, the s_j's representing the seasonal effects with $s_t = s_{t-12}$.

Equation (5.2.5) describes a process with a linear trend and additive seasonal component while equation (5.2.6) describes a linear trend and multiplicative seasonal component.

a) Additive Trend and Seasonals

Simple differencing of the equation (5.2.5) results in

$$\nabla d_t = \beta + s_t - s_{t-1}$$

$$= \beta + s'_t \qquad 5.2.7$$

with $s'_t = s_t - s_{t-1} = s_{t-12} - s_{t-13} = s'_{t-12}$

Thus, the linear trend is removed but a seasonal component s'_t with period 12 still remains. On the other hand the seasonal operator ∇_{12} removes both the linear term in the trend and the seasonal component since

$$\nabla_{12} d_t = 12\beta \qquad 5.2.8$$

The constant term 12β can be removed, if required, by using the simple differencing operator ∇ in conjunction with the seasonal operator. However this is not essential since the stationary A.R.M.A. models can be fitted to series with constant or zero means.

In general, for a polynomial trend of order r, the operator $\nabla^r \nabla_{12}$ will completely remove both trend and seasonal components while the operator $\nabla^{r-1} \nabla_{12}$ will produce a trend-free, deseasonalised series with a non-zero mean.

b) Multiplicative Trend and Seasonals

Although the model (5.2.6) is non-linear in terms of trend and seasonality, Wilson (1973) has shown that a suitable choice of differencing operator will obliterate the deterministic component d_t, thus avoiding the use of non-linear transformations of the form dealt with in Chapter 4.

Simple differencing of the model (5.2.6) gives

$$\nabla d_t = (\alpha + \beta t)(s_t - s_{t-1}) + \beta s_{t-1} \qquad 5.2.9$$

and obviously this does not produce the desired effect. Turning to the seasonal operator ∇_{12} we get

$$\begin{aligned} \nabla_{12} d_t &= 12\beta\, s_{t-12} \\ &= 12\beta\, s_t \end{aligned} \qquad 5.2.10$$

and so the trend component is removed, leaving just a seasonal component Re-employment of the operator ∇_{12} leads to

$$\nabla_{12}^2\, d_t = 0 \qquad 5.2.11$$

In the case of a polynomial trend of order r, the operator ∇_{12}^{r+1} would be the appropriate choice.

It is interesting to note that the operator ∇_{12}^2 will also remove a linear trend and additive seasonal component which would explain why, in Section 4.4, the log-likelihood function for the model based on ∇_{12}^2 was found to be flatter than those of the models based on the operator $\nabla\nabla_{12}$.

Summarising, it would appear that for the additive model (5.2.5) either of the operators ∇_{12} or $\nabla\nabla_{12}$ may be acceptable. It should however be remembered that in the above discussion the error structure has not been considered and so the appropriate choice may be suggested by an examination of the sample autocorrelation functions of $\nabla_{12}X_t$ and $\nabla\nabla_{12}X_t$ following the arguments of Section 5.2.1.

The results for the multiplicative model (5.2.6) are probably more useful than those for the additive model (5.2.5) since the identification of the operator ∇_{12}^2 may prove extremely difficult using the Box-Jenkins approach described in Section 5.2.1. This is especially true for shorter series.

5.3 The Effect of Differencing on the Variance

In addition to inspecting the sample autocorrelation function (as discussed in Section 5.2.1) for various differences of the series X_t, we will demonstrate that it is also useful to keep track of the variance of the differenced series at each stage.

Let us suppose that w_t is a stationary series with variance σ_w^2. Taking first differences of w_t would yield

$$\nabla w_t = w_t - w_{t-1} \qquad 5.3.1$$

and the variance of the differenced series is given by

$$V[\nabla w_t] = 2\sigma_w^2 (1 - \rho_w(1)) \qquad 5.3.2$$

where $\rho_w(k)$ is the k^{th} autocorrelation coefficient of the series w_t.

Generally, for d degrees of differencing, the appropriate variance is

$$V[\nabla^d w_t] = \sigma_w^2 \left[\binom{2d}{d} - 2\rho_w(1)\binom{2d}{d-1} + 2\rho_w(2)\binom{2d}{d-2} - \ldots\ldots \right] \qquad 5.3.3$$

Setting $w'_t = \nabla^d w_t$ and differencing a further D times with respect to the seasonal period results in

$$V[\nabla_{12}^D w'_t] = \sigma_{w'}^2 \left[\binom{2D}{D} - 2\rho_{w'}(12)\binom{2D}{D-1} + 2\rho_{w'}(24)\binom{2D}{D-2} - \ldots\ldots \right] \qquad 5.3.4$$

where $\sigma_{w'}^2$ and $\rho_{w'}(k)$ are respectively the variance and autocorrelation function of the series w'_t.

In identifying the degree of differencing in practice, the above results could be utilised in the following way.

For some particular degree of differencing, the sample variance s_w^2 and the sample autocorrleation function $r_w(k)$ of the differenced series w_t are calculated. Replacing σ_w^2 and $\rho_w(k)$ by s_w^2 and $r_w(k)$ in equation (5.3.3) (similarly for equation (5.3.4)), estimates of the variances for further differences of w_t can be derived. These estimates can be compared with the corresponding sample variances. If w_t is non-stationary, the equations (5.3.3) and (5.3.4) will be invalid and the agreement between the two quantities will be poor. On the other hand, close agreements between the two different estimates would tend to support the assumption that w_t is stationary.

To illustrate the above approach, reference will be made to various differences of the logarithmic transformation of the Company X data. This particular transformation is considered because the initial model in Chapter 3 was identified on the basis of the logged data and also because even when some other transformation is to be entertained the appropriate model must be assumed (from some form of the data) before the transformation parameter can be estimated (see Section 4.4). The transformed series will be denoted by z_t and Table 5.1 shows the estimated variance of $w_t = \nabla^d \nabla^D z_t$ for several values of d and D.

Table 5.1
Estimated Variance of $w_t = \nabla^d \nabla^D z_t$ for various values of d,D

Series	Estimated Variance
z_t	0.0968
∇z_t	0.0200
$\nabla^2 z_t$	0.0259
$\nabla^3 z_t$	0.0806
$\nabla_{12} z_t$	0.0101
$\nabla_{12}^2 z_t$	0.0262
$\nabla\nabla_{12} z_t$	0.0112

We shall first assume that the series $w_t = \nabla z_t$ is stationary. Using equation (5.3.2) with σ_w^2 and $\rho_w(1)$ replaced by their sample estimates, an estimate of the variance of ∇w_t is given by

$$\hat{V}[\nabla w_t] = 2 \times 0.0200 \ (1 - 0.35)$$
$$= 0.0260$$

This agrees closely with the sample variance of $\nabla^2 z_t$ (0.0259, see Table 5.1), suggesting that ∇z_t is stationary with respect to trend. However, let us now examine the effect of differencing ∇z_t further with respect to seasonality. In this case we have

$$\hat{V}[\nabla_{12} w_t] = 2 \times 0.0200 \ (1 - 0.59)$$
$$= 0.0164$$

which does not compare favourably with the sample variance of $\nabla\nabla_{12} z_t$ (0.0112). From this we conclude that ∇z_t is non-stationary with regard to seasonality.

Let us now assume that the series $w_t = \nabla_{12} z_t$ is stationary. On this occasion simple differencing results in

$$\hat{V}[\nabla w_t] = 2 \times 0.0101 \ (1 - 0.45)$$
$$= 0.0111$$

agreeing closely with the sample variance of $\nabla\nabla_{12} z_t$ (0.0112). Differencing over the seasonal period leads to

$$\hat{V}[\nabla_{12} w_t] = 2 \times 0.0101 \ (1 + 0.26)$$
$$= 0.0255$$

Again the agreement with the sample variance of $\nabla_{12}^2 z_t$ (0.0262) is good, so that the variance of $\nabla_{12} z_t$ behaves like that of a stationary series. It is recalled that in Section 5.2.2 we concluded

that the operator ∇_{12} may prove acceptable even though the inspection of the sample autocorrelation function (see Table 3.2) suggested the use of $\nabla\nabla_{12}$. The employment of the operator ∇_{12} as opposed to $\nabla\nabla_{12}$ will be looked at further in Sections 5.4 and 5.5.

5.4 Autocorrelations Introduced by Differencing

In this section we consider once again the model (5.2.3) in which the deterministic component is composed of a linear trend and an additive seasonal effect, i.e. the model

$$X_t = \alpha + \beta t + s_t + n_t \qquad 5.4.1$$

where α, β, s_t and n_t have been defined in Section 5.2.2. Also in Section 5.2.2, we discussed the choice of differencing operator capable of removing the deterministic component from the series X_t. It was found that both the operators ∇_{12} and $\nabla\nabla_{12}$ removed the trend and seasonal components although in the case of the former a constant term still remained. We now look at the effect that the choice of differencing operator has on the error component n_t, for both stationary and non-stationary structures.

The following stationary error structures will be considered:

i) $n_t = a_t$

ii) $n_t = a_t - \theta a_{t-1}$

iii) $n_t = a_t - \Theta a_{t-12}$

where a_t, a_{t-1}, a_{t-2}, ... is a white noise process with variance σ_a^2.

Table 5.2 shows the autocorrelation coefficients for lags 1-12 of the series $\nabla_{12} X_t$ and $\nabla\nabla_{12}X_t$ when the error structure takes the above three forms.

Table 5.2

Theoretical autocorrelation coefficients (lags 1-12) for $\nabla_{12}X_t$ and $\nabla\nabla_{12}X_t$

Lag	$n_t = a_t$		$n_t = a_t - \theta a_{t-1}$		$n_t = a_t - \Theta a_{t-12}$	
	$\nabla_{12}X_t$	$\nabla\nabla_{12}X_t$	$\nabla_{12}X_t$	$\nabla\nabla_{12}X_t$	$\nabla_{12}X_t$	$\nabla\nabla_{12}X_t$
1	0	$-\frac{1}{2}$	$-\frac{\theta}{(1+\theta^2)}$	$-\frac{(1+\theta^2)}{2(1+\theta+\theta^2)}$	0	$-\frac{1}{2}$
2	0	0	0	$\frac{\theta}{2(1+\theta+\theta^2)}$	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0
9	0	0	0	0	0	0
10	0	0	0	$-\frac{\theta}{4(1+\theta+\theta^2)}$	0	0
11	0	$\frac{1}{4}$	$\frac{\theta}{2(1+\theta^2)}$	$\frac{(1+\theta)^2}{4(1+\theta+\theta^2)}$	0	$\frac{(1+\theta)^2}{4(1+\theta+\theta^2)}$
12	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{(1+\theta)^2}{2(1+\theta+\theta^2)}$	$-\frac{(1+\theta)^2}{2(1+\theta+\theta^2)}$

It can be seen from the above table that for the three stationary error structures considered, an A.R.M.A. model based on the operator $\nabla\nabla_{12}$ would include more moving average parameters than one based on the operator ∇_{12}. For example, when $n_t = a_t$, the appropriate model based on $\nabla\nabla_{12}X_t$ would be

$$w_t = \nabla\nabla_{12}X_t = (1 - \theta_1 B)(1 - \theta_1 B^{12})\, a_t \qquad 5.4.2$$

while the model based on $\nabla_{12}X_t$ would be

$$w_t = \nabla_{12}X_t = (1 - \theta_1 B^{12})a_t + C \qquad 5.4.3$$

where C is a constant which is generally easier to estimate than an extra moving average parameter (see Section 7.4).

A similar situation arises when autoregressive error structures are assumed. However, let us now look at the following three simple cases when the errors are non-stationary.

i) $\nabla n_t = a_t$

ii) $\nabla_{12} n_t = a_t$

iii) $\nabla\nabla_{12} n_t = a_t$

The autocorrelation coefficients (lags 1-12) for $\nabla_{12} X_t$ and $\nabla\nabla_{12} X_t$ in these three instances are given in Table 5.3.

Table 5.3. Theoretical autocorrelation coefficients (lags 1-12) for $\nabla_{12} X_t$ and $\nabla\nabla_{12} X_t$

Lag	$\nabla n_t = a_t$		$\nabla_{12} n_t = a_t$		$\nabla\nabla_{12} n_t = a_t$	
	$\nabla_{12} X_t$	$\nabla\nabla_{12} X_t$	$\nabla_{12} X_t$	$\nabla\nabla_{12} X_t$	$\nabla_{12} X_t$	$\nabla\nabla_{12} X_t$
1	11/12	0	0	$-\frac{1}{2}$	NONSTATIONARY	0
2	10/12	0	0	0		0
3	9/12	0	0	0		0
4	8/12	0	0	0		0
5	7/12	0	0	0		0
6	6/12	0	0	0		0
7	5/12	0	0	0		0
8	4/12	0	0	0		0
9	3/12	0	0	0		0
10	2/12	0	0	0		0
11	1/12	0	0	0		0
12	0	$-\frac{1}{2}$	0	0		0

For the non-stationary error structure $\nabla n_t = a_t$, both $\nabla_{12}X_t$ and $\nabla\nabla_{12}X_t$ are stationary processes. From Table 5.3 we see however that $\nabla\nabla_{12}X_t$ possesses a much simpler autocorrelation function than $\nabla_{12}X_t$ and hence a model based on the operator $\nabla\nabla_{12}$ would be easier to identify. In the case of the error structure $\nabla_{12}n_t = a_t$ the operator ∇_{12} would present less difficulties than the operator $\nabla\nabla_{12}$ but when the errors follow the process $\nabla\nabla_{12}n_t = a_t$ the series $\nabla_{12}X_t$ is non-stationary and so $\nabla\nabla_{12}$ must be used.

Thus for the model (5.4.1) it is not possible to draw any general conclusions regarding the appropriate choice of differencing operator. In practice, the structure of the errors n_t would not be known although Box and Newbold (1971) do suggest that in the case of economic models the errors might best be represented by "some stable non-stationary noise model". Even so, as we have seen in the previous paragraph, the best choice of differencing operator still depends to what degree the errors are non-stationary.

5.5 The Influence of the Degree of Differencing on Tolerance Limits

In the previous three sections much emphasis has been placed on the identification of the degree of differencing necessary to induce stationarity in a time series. We now look at a facet of the Box-Jenkins procedure on which the choice of differencing operator can have a considerable effect.

Using the approach described in Sections 2.4.1 and 5.2.1 an A.R.I.M.A. model of the form

$$(1 - \phi B)\ \nabla\nabla_{12}z_t = (1 - \Theta B^{12})a_t \qquad 5.5.1$$

was identified for the Company X data, where $z_t = X_t^{0.23}$ (see Section 4.3). As we have seen in Sections 5.2, 5.3 and 5.4 there is a case for fitting a model to the transformed data using only the seasonal differencing operator ∇_{12}. This being so, the identified model obtained by

inspecting the sample autocorrelation function of $\nabla_{12} z_t$ was found to be

$$\nabla_{12} z_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta B^{12}) a_t + C \qquad 5.5.2$$

where C is a constant and again $z_t = X_t^{0.23}$.

Now it has been shown in Section 2.5 that the approximate $(1 - \alpha) \times 100\%$ probability limits associated with a forecast $\hat{z}_t(\ell)$, made at time t for the future observation $z_{t+\ell}$, are

$$\hat{z}_t(\ell) \pm u_{\alpha/2} \left\{ 1 + \sum_{j=1}^{\ell-1} \psi_j^2 \right\}^{\frac{1}{2}} s_a$$

where $u_{\alpha/2}$ is the appropriate percentage point of the standard normal distribution, s_a^2 is the sample variance of the a_t's and the ψ_j's are arrived at by expressing the relevant model as an infinite moving average, i.e.

$$z_t = \psi_o a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots.$$

and $\psi_o = 1$. Thus the two quantities which affect the width of the probability limits are the ψ weights and the sample standard deviation s_a.

Expressions for the ψ_j's (j = 1,2,3,...,12) for the models (5.5.1) and (5.5.2) are given in Table 5.4.

Table 5.4.

Expressions for the ψ Weights for Models (5.5.1) and (5.5.2)

Model \ j	0	1	2	3	4	5	6	7
(5.5.1)	1	$1+\phi$	$1+\phi+\phi^2$	$1+\phi+..+\phi^3$	$1+\phi+..+\phi^4$	$1+\phi+..+\phi^5$	$1+\phi+..+\phi^6$	$1+\phi+..+\phi^7$
(5.5.2)	1	$-\theta_1$	$-\theta_2$	0	0	0	0	0

Table 5.4 (Cont'd.)

Model \ j	8	9	10	11	12
(5.5.1)	$1+\phi+..+\phi^8$	$1+\phi+..+\phi^9$	$1+\phi+..+\phi^{10}$	$1+\phi+..+\phi^{11}$	$1-\theta+1+\phi+..+\phi^{12}$
(5.5.2)	0	0	0	0	$1-\theta$

Substituting the parameter estimates obtained for both models into the above expressions, numerical values for the ψ_j's were computed. Using these estimated values the quantity

$$W(\ell) = \left\{ 1 + \sum_{j=1}^{\ell-1} \psi_j^2 \right\}^{\frac{1}{2}} \qquad 5.5.3$$

was evaluated for $\ell = 1,2,3,\ldots,24$, for both the models (5.5.1) and (5.5.2). $W(\ell)$ can be thought of as the ratio of the standard deviations of the lead time ℓ forecast errors and the lead time 1 forecast errors. Thus $W(1) = 1$ and $W(\ell) \geq 1$ for $\ell > 1$.

$W(\ell)$ ($\ell = 1,2,3,\ldots,24$) is plotted as a function of ℓ in Figure 5.1. It can be seen that while for model (5.5.1) $W(\ell)$ increases steadily (almost linearly) with ℓ, reaching a value of 3.61 for $\ell = 24$, for model (5.5.2) $W(\ell)$ takes a value of 1.23 when $\ell = 2$ and thereafter remains constant apart from a slight increase at lead times 13, 14 and 15. The implication of this is that in the case of model (5.5.1) much more confidence can be placed in the lead time 1 forecasts than in the forecasts for higher lead times while for model (5.5.2) the forecasts for higher lead times can be expected to be almost as accurate as those for lead time 1.

So far we have been examining the tolerance limits, for various lead times, within each model. Another, more important, problem is to compare the tolerance limits resulting from the model (5.5.1) with those from the model (5.5.2). To do this of course we have to take

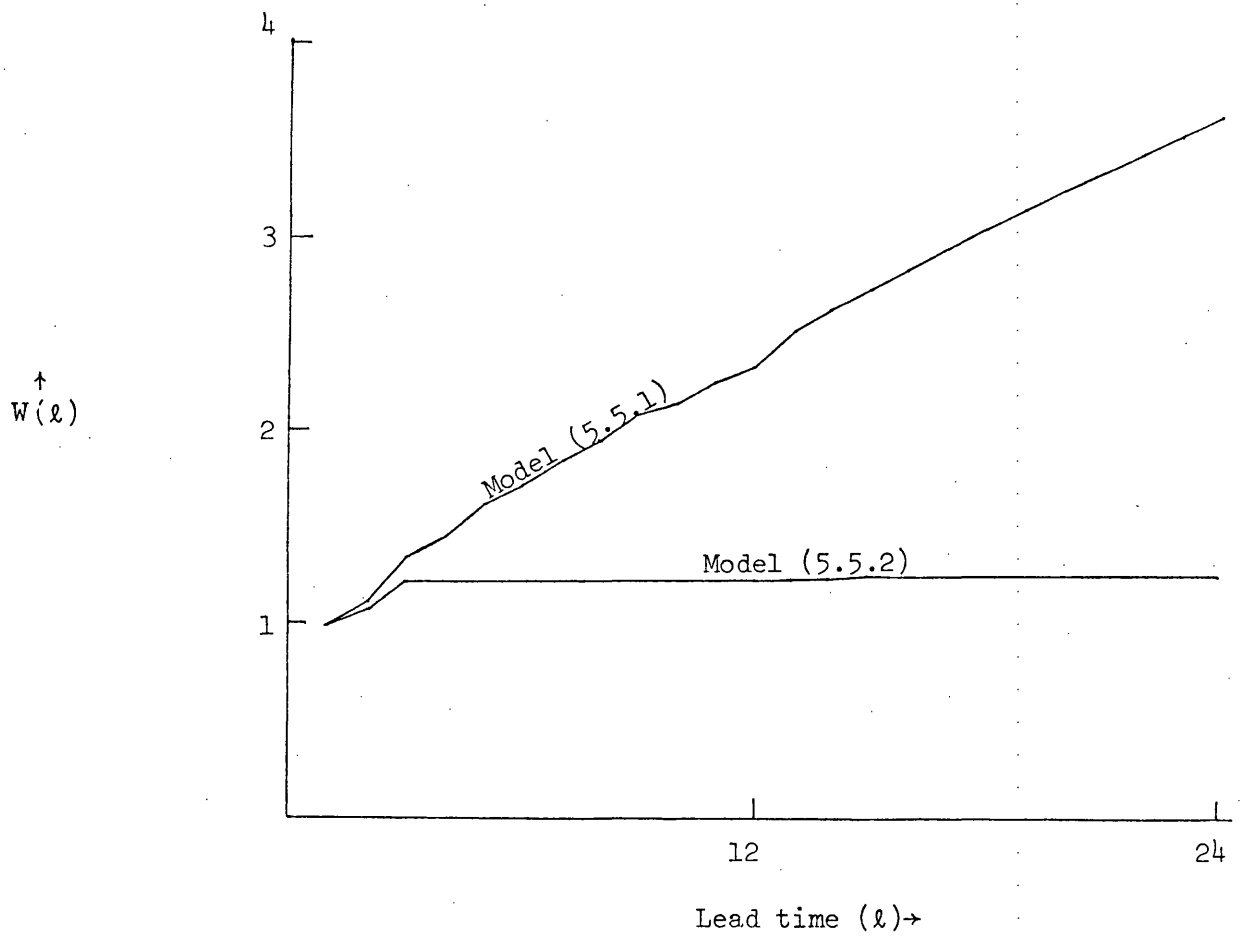


FIGURE 5.1

W(ℓ) for models (5.5.1) and (5.5.2)

into account the estimated standard deviations of the lead time l forecast errors (or the a_t's). These standard deviations will be denoted by s_{a_1} and s_{a_2} for the models (5.5.1) and (5.5.2) respectively. Further, $W_1(\ell)$ and $W_2(\ell)$ are defined by equation (5.5.3) with the subscripts referring to the appropriate model. A measure of the relative accuracy of the forecasts derived from the two models will then be given by

$$R(\ell) = \frac{W_1(\ell)\, s_{a_1}}{W_2(\ell)\, s_{a_2}} \qquad 5.5.4$$

When $\ell = 1$, $W_1(1) = W_2(1)$ and so

$$R(1) = \frac{s_{a_1}}{s_{a_2}}$$

i.e. $R(1)$ is simply the ratio of the estimated standard deviations of the a_t's for each model.

Another interesting case is if $s_{a_1} = s_{a_2}$. Then we have

$$R(\ell) = \frac{W_1(\ell)}{W_2(\ell)}$$

and values of $R(\ell)$ can be obtained quite easily from Figure 5.1. These values (for $\ell = 1,2,3,\ldots,24$) are shown in Table 5.5.

Table 5.5. Values of $R(\ell)$, $\ell = 1,2,3,\ldots,24$, when $s_{a_1} = s_{a_2}$

Lead Times	1-12	1	1.03	1.10	1.20	1.33	1.42	1.52	1.61	1.70	1.77	1.85	1.93
Lead Times	13-24	2.03	2.11	2.19	2.28	2.37	2.45	2.54	2.60	2.67	2.74	2.82	2.89

Thus, if $s_{a_1} = s_{a_2}$ the probability limits associated with the point forecasts derived from the model (5.5.1) would always be wider than those for the model (5.5.2), after lead time 1.

In our particular example we found that

$$\frac{s_{a_1}}{s_{a_2}} = \frac{0.1259}{0.1226} = 1.03$$

so that the values for $R(\ell)$ can be obtained by multiplying throughout in Table 5.5 by a factor 1.03. Thus the probability limits resulting from use of the model (5.5.1) were in fact wider than those for model (5.5.2), for all lead times.

Although it is desirable to obtain probability limits (for all lead times) which are as narrow as possible, it should be remembered that the way in which Box and Jenkins determine these limits assumes that the true model has been fitted. In the case of the Company X data, if the model (5.5.1) was known to represent the true underlying process then use of the model (5.5.2) would give far too much confidence to the computed point forecasts for higher lead times. On the other hand, if the true process was described by the model (5.5.2) then the model (5.5.1) would lead to unnecessarily wide tolerance limits. Have we a means therefore of deciding which of the models (5.5.1) and (5.5.2) is nearer the true underlying process? The diagnostic checks described in Section 2.4.3 did not reveal any inadequacies in either of the models, but of course these checks are primarily concerned with the autocorrelation properties of the residuals and not with the forecasting performance of the model. In order to look more closely at the models from a forecasting point of view the following approach, similar to that used by Reid (1969), was adopted.

Both models were re-fitted to the first 60 observations of the

transformed* Company X data, z_t, and forecasts were generated over the remaining 17 observations. The mean squared error function M.S.E.(ℓ) was calculated for lead times 1-12, for each model. It should be stressed that in the case of model (5.5.1) z_t is considered to be defined from t = -12, -11, -10, ..., 64 while for model (5.5.2) z_t is defined from t = -11, -10, -9, ..., 65 so that the mean squared error function is defined by

$$\text{M.S.E.}(\ell) = \begin{cases} \dfrac{1}{17-\ell+1} \displaystyle\sum_{t=47+\ell}^{64} \left[z_t - \hat{z}_{t-\ell}(\ell)\right]^2 \\ \qquad \text{for model (5.5.1)} \\ \dfrac{1}{17-\ell+1} \displaystyle\sum_{t=48+\ell}^{65} \left[z_t - \hat{z}_{t-\ell}(\ell)\right]^2 \\ \qquad \text{for model (5.5.2)} \end{cases} \qquad 5.5.5$$

Now following the Box-Jenkins procedure, the mean squared error for lead time ℓ would be related to the mean squared error for lead time 1 by the equation

$$\text{M.S.E.}(\ell) = \left[1 + \sum_{j=1}^{\ell-1} \psi_j^2\right] \text{M.S.E.}(1) \qquad 5.5.6$$

$$\text{for } \ell = 2,3,4,\ldots,12.$$

For convenience, the mean squared error defined by equation (5.5.5) will be referred to as the sample mean squared error (S.M.S.E.(ℓ)) and that defined by equation (5.5.6) as the theoretical mean squared error (T.M.S.E.(ℓ)). If S.M.S.E.(ℓ) and T.M.S.E.(ℓ) are in close agreement for all ℓ then the model is under examination can be considered to provide a good approximation to the true underlying process. However, if S.M.S.E.(ℓ) and T.M.S.E.(ℓ) compare unfavourably then doubt can be expressed about the model in question.

---

*The transformation parameter was also re-estimated.

The mean squared error functions S.M.S.E.(ℓ) and T.M.S.E.(ℓ) are tabulated below for both the models (5.5.1) and (5.5.2).

Table 5.6. S.M.S.E.(ℓ) and T.M.S.E.(ℓ) (ℓ=1,2,3,...,12) for Models (5.5.1) and (5.5.2)

Lead Time	Model (5.5.1)		Model (5.5.2)	
	S.M.S.E.(ℓ)	T.M.S.E.(ℓ)	S.M.S.E.(ℓ)	T.M.S.E.(ℓ)
1	0.212	0.212	0.199	0.199
2	0.245	0.297	0.201	0.255
3	0.356	0.422	0.239	0.329
4	0.377	0.533	0.239	0.329
5	0.363	0.645	0.236	0.329
6	0.308	0.758	0.231	0.329
7	0.304	0.870	0.252	0.329
8	0.237	0.983	0.272	0.329
9	0.210	1.095	0.297	0.329
10	0.338	1.207	0.320	0.329
11	0.279	1.320	0.360	0.329
12	0.541	1.432	0.415	0.329

From Table 5.6 it can be seen that after lead time 3 there is a closer agreement between S.M.S.E.(ℓ) and T.M.S.E.(ℓ) for model (5.5.2) than for model (5.5.1). This would suggest that the model based on the single differencing operator ∇_{12} is nearer the true model. However, the sample mean squared errors are based on so few observations (17 for S.M.S.E.(1), 6 for S.M.S.E.(12)) that to reject the model (5.5.1) (particularly since it was the one suggested by inspecting the sample autocorrelation function) would be somewhat unwise.

For larger samples, the above procedure would probably prove quite

successful for deciding between A.R.I.M.A. models which fit the data almost equally well. Granger (1973) has proposed a technique by which simultaneous confidence limits can be placed on T.M.S.E.(ℓ) (or equivalently the sequence of error variances) and it can then be observed whether the S.M.S.E.(ℓ) fall within these limits. The determination of these confidence limits involves the assumption that the forecasting period is long compared with the lead time being forecasted. In our case this assumption would only be valid for short lead times and so the technique was not applied.

Our experience in this section does tend to highlight a problem encountered in Chapter 3 and mentioned in Chatfield and Prothero (1973a), namely that in employing the Box-Jenkins procedure it is often possible to find several A.R.I.M.A. models which fit the data equally well yet generate quite different point forecasts and/or tolerance limits. For short series, it may not be possible to decide which model represents the closest approximation to the true generating process.

5.6 Summary

The selection of the differencing operator necessary to reduce a non-stationary series to a stationary process need not always be based entirely on an inspection of the sample autocorrelation function, as described in Section 5.2.1. In Sections 5.2 and 5.3 two complementary approaches have been proposed.

When analysing seasonal data which have a relatively small random variation, the series in question can be resolved into trend and seasonal components and the appropriate differencing operator will be one which will remove the trend and seasonality. The same approach may not be applicable in cases where the random variation is more dominant since it may be more difficult to recognise the trend and seasonal components and the choice of differencing operator may depend on the structure of the errors. No such restrictions need be placed on the use of the

technique based on the theoretical behaviour of the variance of a stationary series, described in Section 5.3. This technique can be applied to both seasonal and non-seasonal series.

The importance of identifying the correct degree of differencing has been demonstrated in Section 5.5, where a model based on the single differencing operator ∇_{12} gave rise to much narrower tolerance limits (for higher lead times) than those obtained from a model based on the double operator $\nabla\nabla_{12}$. The models fitted the data almost equally well and the diagnostic checks did not reveal any serious inadequacies in either. Thus even though with respect to the one step ahead forecasts it did not really matter which model was employed (see Box and Jenkins (1973)), for higher lead times the consequences of fitting an inadequate model become more serious.

CHAPTER 6

COMPUTATION OF THE UNCONDITIONAL SUM OF SQUARES

6.1 Introduction

The two most commonly used procedures for estimating the parameters in autoregressive-moving average models are the graphical technique outlined in Section 2.4.2 and the non-linear least squares approach described by Box and Jenkins (1970, pages 231-242). Both methods involve the computation of the unconditional sum of squares defined in Section 2.4.2. In computing this sum of squares it is possible to perform more than just the one iterative cycle which is illustrated for example in Table 3.3. Although Box and Jenkins (1970, pages 218, 318) refer to this possibility, they do not state any conditions under which further iterations may be necessary, apart from mentioning that in practice "a second iterative cycle would almost never be needed." On the contrary, we shall see in Chapter 10 that for 4 of the 5 series analysed in this thesis, one iterative cycle was not sufficient and so clearly there is need to discuss the problem in detail. This chapter is therefore concerned with situations in which more than one iterative cycle is necessary in order for the sum of squares to converge.

The steps included in the computation of the unconditional sum of squares when several iterative cycles have to be performed will be outlined in Section 6.2. The additional steps involved in this process will suggest cases when more than one iteration should be entertained.

In Section 6.3 the procedure set out in Section 6.2 will be applied to fitting models to the Company X data and the results will be compared with those obtained when only one iterative cycle is

employed.

The conclusions of the work described in this chapter will be stated in Section 6.4.

6.2 The Full Procedure for Calculating the Unconditional Sum of Squares

The series w_t $(t=1,2,3,\ldots,n)$ is assumed to be stationary and described by the multiplicative seasonal A.R.M.A. model

$$\phi_p(B)\Phi_P(B^{12})w_t = \theta_q(B)\Theta_Q(B^{12})a_t \qquad 6.2.1$$

where for convenience the seasonal cycle is considered to have a period 12. It is further assumed, without loss of generality that w_t has a zero mean. The operators $\phi_p(B)$, $\Phi_P(B^{12})$, $\theta_q(B)$ and $\Theta_Q(B^{12})$ have the usual meaning (see Section 2.6) and the corresponding vectors of parameters will be denoted by $\underline{\phi}$, $\underline{\Phi}$, $\underline{\theta}$ and $\underline{\Theta}$.

As stated in Section 2.4.2, the w_t's generated by the model (6.2.1) could equally have been generated by the model

$$\phi_p(F)\Phi_P(F^{12})w_t = \theta_q(F)\Theta_Q(F^{12})e_t \qquad 6.2.2$$

where F is the forward shift operator and e_t is a white noise process possessing variance σ_e^2 $(= \sigma_a^2)$.

In order to describe the computation of the unconditional sum of squares for the model (6.2.1) as clearly as possible, a step-by-step approach will be adopted.

STEP 1

Set

$$[e_t] = 0 \qquad \text{for } t>n-(p+12P)$$

where [ ] is used to denote the expectation conditional on $\underline{\phi}$, $\underline{\Phi}$, $\underline{\theta}$,

$\underline{\theta}$ and $w_1, w_2, w_3, \ldots, w_n$.

STEP 2

Starting with t=n-(p+12P),calculate the $[e_t]$'s recursively, in reverse order until t = 1, using the equation

$$\phi_p(F)\Phi_P(F^{12})[w_t] = \theta_q(F)\theta_Q(F^{12})[e_t] \qquad 6.2.3$$

where $[w_t] = w_t$, for t=1,2,3,...,n.

STEP 3

Set

$$[e_t] = 0 \qquad \text{for } t \leq 0$$

and generate the backward forecasts $[w_o]$, $[w_{-1}]$, $[w_{-2}]$,...,$[w_{-K_1+1}]$, using equation (6.2.3.) The integer K_1 is chosen so that the $[w_t]$'s are negligible for $t \leq -K_1$.

STEP 4

Set

$$[a_t] = 0 \qquad \text{for } t \leq -K_1$$

and calculate the $[a_t]$'s for $t=-K_1+1, -K_1+2, -K_1+3, \ldots, n$, recursively, using the equation

$$\phi_p(B)\Phi_P(B^{12})[w_t] = \theta_q(B)\theta_Q(B^{12})[a_t] \qquad 6.2.4$$

STEP 5

Sum the squares of the $[a_t]$'s from $t=-K_1+1, -K_1+2, -K_1+3, \ldots, n$, to obtain the unconditional sum of squares $S(\underline{\phi}, \underline{\Phi}, \underline{\theta}, \underline{\Theta})$ i.e.

$$S(\underline{\phi}, \underline{\Phi}, \underline{\theta}, \underline{\Theta}) = \sum_{t=-K_1+1}^{n} [a_t]^2 \qquad 6.2.5$$

The steps 1-5 constitute one complete iterative cycle and in

all the estimation problems encountered thus far only one such iteration has been employed. In order to perform more than one iterative cycle the following additional steps are required.

STEP 6

Obtain the forward forecasts $[w_t]$ for $t=n+1,n+2,n+3,\ldots,n+K_2-1$, using equation (6.2.4) where the $[a_t]$'s $(t=-K_1+1,-K_1+2,-K_1+3,\ldots,n)$ are generated in step 4 and

$$[a_t] = 0 \qquad \text{for } t>n$$

The integer K_2 is chosen so that $[w_t]$ is negligible for $t\geq n+K_2$.

STEP 7

Set

$$[e_t] = 0 \qquad \text{for } t\geq n+K_2$$

STEP 8

Starting with $t=n+K_2-1$, calculate the $[e_t]$'s recursively, in reverse order until $t=1$, using equation (6.2.3) where the $[w_t]$'s for $t>n$ are obtained in step 6 and $[w_t] = w_t$ for $t=1,2,3,\ldots,n$.

STEP 9

A new value for $S(\underline{\phi},\underline{\Phi},\underline{\theta},\underline{\Theta})$ is computed via steps 3, 4 and 5.

More iterative cycles can be performed by following the steps 6, 7, 8 and 9 until the sum of squares is judged to converge.

In order to seek situations where the additional steps 6, 7, 8 and 9 may prove necessary we will now consider some special cases of the seasonal model (6.2.1). Let us first examine the purely autoregressive model

$$\phi_p(B)\Phi_P(B^{12})w_t = a_t \qquad 6.2.6$$

In step 3 the backward forecasts $[w_o], [w_{-1}], [w_{-2}], \ldots, [w_{-K_1+1}]$ would be generated using the equation

$$\phi_p(F)\Phi_P(F^{12})[w_t] = [e_t] \qquad 6.2.7$$

but $[e_t] = 0$ for $t \leq 0$ so that equation (6.2.7) becomes

$$\phi_p(F)\Phi_P(F^{12})[w_t] = 0 \qquad 6.2.8$$

for $t \leq 0$. Thus, the backward forecasts do not depend on the $[e_t]$'s for $t>0$ and so the steps 6, 7, 8 and 9 which affect the $[e_t]$'s (for $t>0$) would not affect $[w_o], [w_{-1}], [w_{-2}]$ etc. Hence a single iteration will always prove sufficient when computing the unconditional sum of squares for a pure autoregressive process.

It should be noted that equation (6.2.8) has a solution of the form

$$[w_t] = A_1\left(\frac{1}{G_1}\right)^t + A_2\left(\frac{1}{G_2}\right)^t + \ldots + A_{p+12P}\left(\frac{1}{G_{p+12P}}\right)^t \qquad 6.2.9$$

for $t \leq 0$, where the G_i's are assumed to be distinct and defined by $\phi_p(B)\Phi_P(B^{12}) = \prod_{i=1}^{p+12P}(1 - G_iB)$ and the A_i's ($i = 1,2,3,\ldots,p+12P$) are constants. Thus, if one of the G_i's is close to ± 1 (i.e. the process defined by equation (6.2.6) approaches non-stationarity), $[w_t]$ will not die out quickly as t decreases and a large value for K_1 will result. This will tend to make the estimation procedure rather lengthy.

We now consider some purely moving average processes, beginning with the model

$$w_t = (1 - \theta B)a_t \qquad 6.2.10$$

The backward forecast $[w_o]$ generated in step 3 is given by

$$[w_o] = -\theta [e_1]$$

$$= -\theta\{[w_1] + \theta[w_2] + \theta^2[w_3] + \ldots + \theta^{n-1}[w_n]$$

$$+ \theta^n[e_{n+1}]\} \qquad 6.2.11$$

with $[w_t] = w_t$ for $t=1,2,3,\ldots,n$. In performing the first iterative cycle $[e_{n+1}]$ is set equal to zero while for further iterations $[e_{n+1}]$ is set equal to the forecast $[w_{n+1}]$ computed in step 6. Provided θ is not too close to ± 1, the coefficient θ^{n+1} will be negligibly small for moderate length series and so $[w_o]$ would be unaffected by the starting value $[e_{n+1}]$. Thus, only one iteration would be necessary. On the other hand if θ is very close to ± 1 then for short series a second iterative cycle may have to be entertained.

Let us look at the seasonal moving average model

$$w_t = (1 - \theta B^{12})a_t \qquad 6.2.12$$

and it is assumed for convenience that the seasonal cycle is repeated m times i.e. $n = 12m$. The backward forecasts $[w_t]$ $(t=0,-1,-2,\ldots,-11)$ can be expressed as

$$[w_t] = -\theta[e_{t+12}]$$

$$= -\theta\{[w_{t+12}] + \theta[w_{t+24}] + \theta^2[w_{t+36}] + \ldots$$

$$+ \theta^{m-1}[w_{t+12m}] + \theta^m[e_{t+12(m+1)}]\} \qquad 6.2.13$$

For the first iteration, $[e_{t+12(m+1)}] = 0$ $(t=0,-1,-2,\ldots,-11)$, while for ensuing iterations $[e_{t+12(m+1)}] = [w_{t+12(m+1)}]$ where the forecasts $[w_{t+12(m+1)}]$ for $t = 0,-1,-2,\ldots,-11$ are generated in step 6. On this occasion, the weight given to the starting values $[e_{t+12(m+1)}]$ $(t = 0,-1,-2,\ldots,-11)$ is governed by m+1 rather than n+1 and so for values of θ close to ± 1 it may be necessary to perform several iterative cycles in computing the unconditional sum of squares, even for series of a moderate length. Once again, for small values of θ one iteration should suffice.

Generally speaking, when computing the unconditional sum of squares for a moving average process one should always be aware of the possibility that more than one iteration may be required. Further iterative cycles are most likely to be needed in situations where the process in question approaches non-invertibility, especially with respect to seasonality.

For autoregressive processes, we noted that the back forecasts $[w_t]$ for $t \leq 0$ generated in step 3 may take a long time to die out. In the case of moving average models this is not so since for the model

$$w_t = \theta_q(B)\Theta_Q(B^{12})a_t \qquad 6.2.14$$

the backward forecasts are given by

$$[w_t] = \theta_q(F)\Theta_Q(F^{12})[e_t] \qquad 6.2.15$$

for $t \leq 0$, where $[e_t] = 0$ for $t \leq 0$ and so

$$[w_t] = 0 \qquad \text{for } t \leq -(q+12Q).$$

Finally, the conditions under which more than one iterative

cycle may be necessary for mixed autoregressive-moving average models are precisely those for the pure moving average models since autoregressive models never require more than one iteration.

6.3 Practical Examples

We first consider the effect of employing only one iterative cycle in computing the unconditional sum of squares, as opposed to iterating until convergence is achieved, when a model of the form

$$(1 - \phi B)\, \nabla\nabla_{12} X_t^{\lambda} = (1 - \theta B^{12}) a_t \qquad 6.3.1$$

is assumed to describe the Company X data. In Chapter 4, the estimates of the parameters in the model (6.3.1) were obtained using the graphical technique described in Section 2.4.2, only one iterative cycle being employed in computing the unconditional sum of squares $S(\phi,\theta)$. These estimates were found to be

$$\hat{\phi} = -0.50, \quad \hat{\theta} = 0.80, \quad \hat{\lambda} = 0.23$$

and $S(-0.50, 0.80) = 1.015$.

Using the same estimate for λ, the autoregressive-moving average parameters in model (6.3.1) were re-estimated by the graphical technique, on this occasion iterations were performed until the unconditional sum of squares was judged to converge. This resulted in

$$\hat{\phi} = -0.48, \quad \hat{\theta} = 0.97$$

and $S(-0.48, 0.97) = 0.854$.

In employing the graphical estimation technique, $S(\phi,\theta)$ was evaluated over a grid of values for ϕ and θ. For smaller values of θ only one iterative cycle was necessary in computing $S(\phi,\theta)$ but as θ became larger more iterations were required to achieve convergence. The behaviour of the minimum sum of squares $S(-0.48, 0.97)$ is shown in Table 6.1.

Table 6.1 Sum of squares $S(-0.48, 0.97)$ for 8 iterations

Iteration	$S(-0.48, 0.97)$	Iteration	$S(-0.48, 0.97)$
1	1.661	5	0.883
2	1.209	6	0.865
3	1.009	7	0.858
4	0.922	8	0.854

$S(-0.48, 0.97)$ was judged to have converged after 8 iterations, although a case could be made for stopping after fewer iterations. It is of interest to note that the minimum sum of squares after one iteration, $S(-0.50, 0.80)$, converged after 2 iterations to a value of 0.973 which lies outside the 95% confidence region for the sum of squares associated with the true parameters.

In addition to reducing the minimum unconditional sum of squares from 1.015 to 0.854, the employment of further iterative cycles caused the point estimate of θ to be changed quite considerably from 0.80 to 0.97. On the contrary, the point estimate of ϕ was virtually unaffected.

Thus far we have considered the re-estimation of the parameters ϕ and θ in the model (6.3.1) using the value of the transformation parameter λ estimated on the basis of the unconditional sum of squares computed using just one iterative cycle. We now examine the re-estimation of the parameter λ.

Figure 6.1 shows the log-likelihood function for the model (6.3.1) when the computation of the unconditional sum of squares was obtained using (a) one iteration, and (b) the number of iterations necessary to achieve convergence. It can be seen that the point estimate of the transformation parameter remains at $\hat{\lambda} = 0.23$ and the respective curves are almost parallel near this maximum, although as λ approaches 1 the lines become closer. This latter characteristic is explained by the fact that as λ increases the point estimate for θ becomes smaller and so less iterations are required in order for the sum of squares to converge.

The fitted models

$$(1 + 0.50B)\ \nabla\nabla_{12} X_t^{0.23} = (1 - 0.80B^{12})a_t \qquad 6.3.2$$

and

$$(1 + 0.48B)\ \nabla\nabla_{12} X_t^{0.23} = (1 - 0.97B^{12})a_t \qquad 6.3.3$$

were used to generate forecasts from May 1971 for lead times 1 to 12. The point forecasts and their associated tolerance limits are given in Table 6.2.

Even though there is quite a large change in the estimates of the seasonal moving average parameter in the fitted models (6.3.2) and (6.3.3), it turns out that the resulting point forecasts actually differ only slightly. Also, while the fitted model (6.3.3)reduces the width of the tolerance limits for all the lead times considered, this reduction is less than 10% in each case. These results tend to support the view of Box and Jenkins (1970, page 308) that the "forecasting procedure is robust to moderate changes in the values of the parameters".

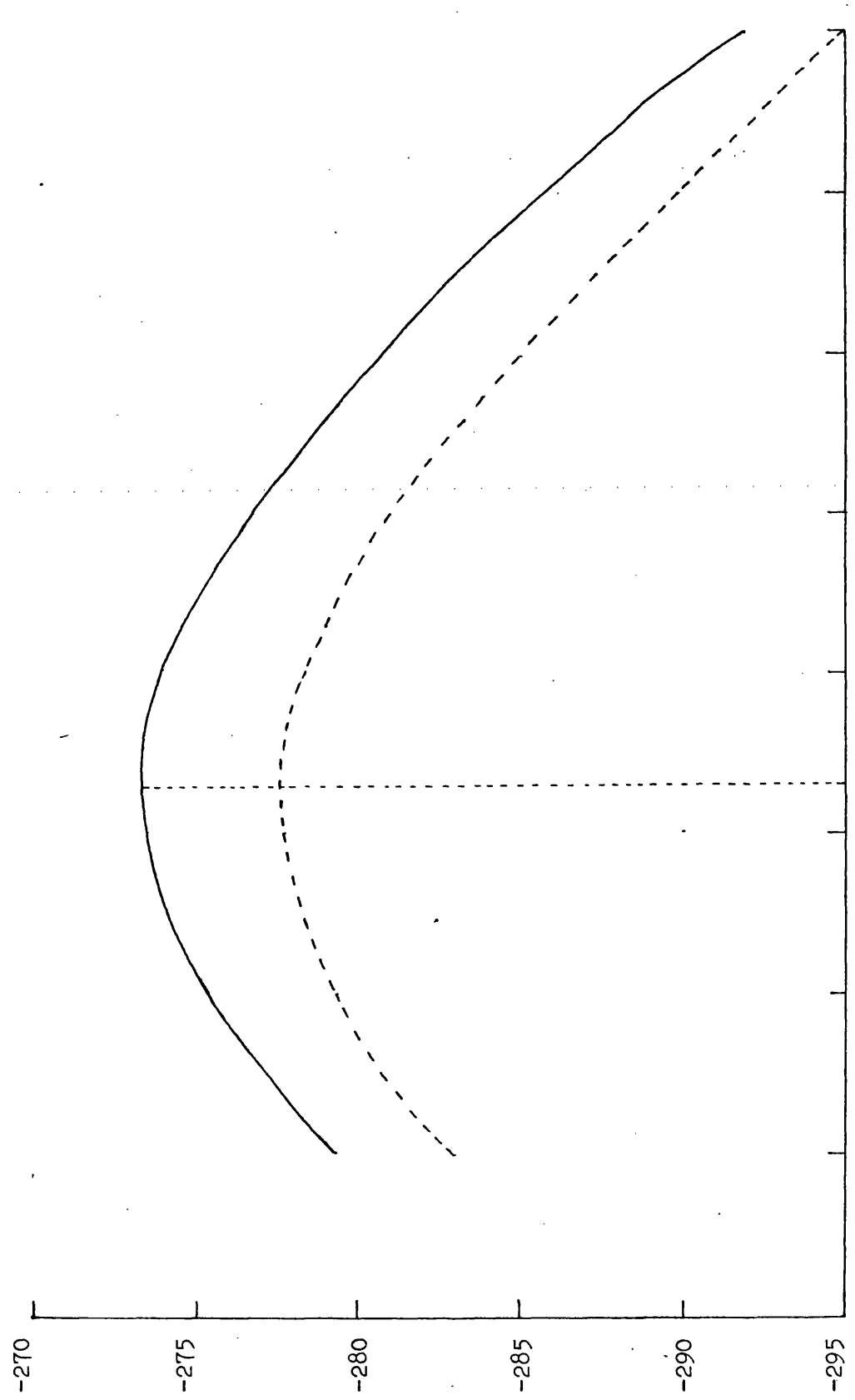


FIGURE 6.1

Log likelihood function for model $(1-\phi B)\nabla\nabla_{12} X_t^{\lambda} = (1-\theta B^{12})a_t$

(a) using one iteration

(b) using several iterations

Table 6.2 Forecasts made at May 1971 for lead times 1 - 12

Lead Time	Fitted Model (6.3.2)	Fitted Model (6.3.3)
1	275 ± 83	282 ± 78
2	399 ± 125	406 ± 116
3	493 ± 176	508 ± 166
4	734 ± 263	742 ± 246
5	929 ± 350	936 ± 323
6	980 ± 393	987 ± 365
7	751 ± 344	754 ± 318
8	742 ± 363	746 ± 335
9	482 ± 276	495 ± 259
10	416 ± 260	426 ± 244
11	352 ± 240	356 ± 222
12	310 ± 228	317 ± 212

The model (6.3.1) has been fitted to the first 60 observations of the Company X data by Wilson (1973). The parameter estimates were derived from the unconditional sum of squares based on a single iteration and the fitted model was

$$(1 + 0.37B)\ \nabla\nabla_{12} X_t^{0.34} = (1 - 0.79B^{12})a_t \qquad 6.3.4$$

For completeness, the parameters ϕ and Θ have been re-estimated by the full procedure described in Section 6.2. This resulted in

$$(1 + 0.37B)\ \nabla\nabla_{12} X_t^{0.34} = (1 - 0.97B^{12})a_t \qquad 6.3.5$$

and forecasts (in terms of the transformed variable) were generated over the remaining 17 observations, for lead times 1 to 12, using both the fitted models (6.3.4) and (6.3.5). The forecasting

performances of the two fitted models were assessed by computing the mean squared forecast errors in each case. These quantities are shown in Table 6.3.

From Table 6.3 it can be seen that the fitted model (6.3.4) produced the smaller mean squared errors for all lead times, after lead time 1. However, as for the point forecasts quoted in Table 6.2, the agreement between the two fitted models is very close. Thus, in our case study, it did not seem to matter greatly from a forecasting point of view whether the unconditional sum of squares was computed using one iterative cycle or by iterating until convergence was achieved.

Table 6.3 Mean squared forecast errors, lead times 1 - 12

Lead Time	Fitted Model (6.3.4)	Fitted Model (6.3.5)
1	0.212	0.205
2	0.245	0.252
3	0.356	0.357
4	0.377	0.395
5	0.363	0.384
6	0.308	0.323
7	0.304	0.309
8	0.237	0.242
9	0.210	0.205
10	0.338	0.349
11	0.279	0.316
12	0.541	0.561

In order to further illustrate some of the points made in Section 6.2, we will briefly discuss the fitting of three other

models to the transformed Company X data. The three models examined were of the same form as the models (B), (C) and (D) introduced in Section 3.7, i.e.

$$(1 - \phi B)(1 - \Phi B^{12})w_t = a_t \qquad 6.3.6$$

$$(1 - \Phi B^{12})w_t = (1 - \theta B)a_t \qquad 6.3.7$$

$$w_t = (1 - \theta B)(1 - \Theta B^{12})a_t \qquad 6.3.8$$

The parameters in each model were estimated using the graphical technique in which the unconditional sum of squares was based on (a) one iterative cycle, and (b) iterating until convergence was achieved.

In the case of the purely autoregressive model (6.3.6), the employment of more than one iterative cycle had no effect whatsoever, for the reasons given in Section 6.2. Again for the model (6.3.7) the use of more than one iteration did not change the value of the sum of squares. The reason for this was that the value of $\theta(= 0.49)$ was far enough away from ± 1. For the model (6.3.8) however, 8 iterations were necessary before the sum of squares was judged to have converged and as for model (6.3.1) the moving average seasonal parameter was estimated to be 0.97.

6.4 Conclusions

Most techniques used for estimating the parameters in autoregressive-moving average time series models involve the computation of the unconditional sum of squares. In evaluating this sum of squares it is possible to perform more than one iterative cycle by employing the procedure described in Section 6.2. This chapter has been concerned with situations in which

several iterations are required before the sum of squares converges.

For purely autoregressive processes one iteration will always suffice but when moving average parameters are involved this is not always the case. In particular for short series it may be necessary to employ more than one iteration when the process in question approaches non-invertibility. This is also true for moderate length series when the moving average operator includes a factor $(1 - GB^{12})$ where G is close to ± 1.

In situations where more than one iteration is appropriate, the resulting estimates of the moving average parameters can differ quite considerably from the estimates obtained when only one iteration is employed. However from our experience with the Company X data such an occurrence may not greatly affect the forecasting performance of the model in question.

CHAPTER 7

SERIES WHICH INCLUDE DETERMINISTIC COMPONENTS

7.1 Introduction

When a time series includes for example a polynomial trend component, Box and Jenkins (1973) say that a forecasting method which involves the fitting of a polynomial regression curve is a special case of the A.R.I.M.A. model. While in theory this is often the case, in practice the fitting of a particular A.R.I.M.A. model in such circumstances can present certain difficulties. In this chapter we will discuss some of the problems encountered in applying the Box-Jenkins procedure to series which include deterministic components and it will be suggested that traditional regression techniques can often be a more practical proposition.

In Section 7.2 we will show that when the Box-Jenkins approach of differencing to produce stationarity is applied to a series which includes deterministic components then under certain conditions the differenced series will be described theoretically by a non-invertible A.R.I.M.A. model. When this situation occurs the Box-Jenkins procedure would involve fitting a model in which the moving average parameters are close to the boundary of the non-invertibility region. Hence, from the results of Chapter 6, the estimation of the moving average parameters may be rather tedious, in which case it would be desirable to employ a more convenient forecasting technique. In Section 7.3 a series based on a deterministic linear trend will be generated and the performance of the Box-Jenkins procedure will be compared with that of a simple regression technique. A further characteristic of the Box-Jenkins procedure when applied to series which include deterministic

components will be examined in Section 7.4 and the conclusions to be drawn from this chapter will be stated in Section 7.5.

7.2 Non-invertibility in A.R.M.A. Models

The reasons for imposing the invertibility condition on A.R.M.A. models will not be given in this section. Instead, reference can be made to Box and Jenkins (1970), Kendall (1971) and Chatfield and Prothero (1973b). The purpose of this section is to discover the kind of series which when differenced gives rise to a non-invertible A.R.M.A. model.

Let us suppose that a series X_t can be described by the additive model discussed in Section 5.2.2. viz.

$$X_t = m_t + s_t + n_t \qquad 7.2.1$$

where m_t represents a deterministic trend, s_t is the seasonal variation (again deterministic) with $s_t = s_{t-12}$ and n_t is the error term. Now if m_t consists of a polynomial of degree r, then we showed in Section 5.2.2 that the differencing operator $\nabla^r \nabla_{12}$ would completely remove the deterministic components m_t and s_t from the series X_t. Thus setting $w_t = \nabla^r \nabla_{12} X_t$ we have

$$w_t = \nabla^r \nabla_{12} n_t = (1-B)^r(1-B^{12})n_t \qquad 7.2.2$$

So far no assumptions have been made about the structure of the errors n_t. Let us suppose therefore that these errors can be described by the general multiplicative A.R.I.M.A. model

$$\phi_1(B)\Phi_1(B^{12})\, \nabla^{d1}\nabla_{12}^{D1}\, n_t = \theta_1(B)\Theta_1(B^{12})\, a_t \qquad 7.2.3$$

where $\phi_1(B)$, $\Phi_1(B^{12})$, $\theta_1(B)$ and $\Theta_1(B^{12})$ are of order p_1, $12P_1$, q_1 and $12Q_1$ respectively.

The model (7.2.2.) thus becomes

$$\phi_1(B)\Phi_1(B^{12})\ \nabla^{d_1}\nabla_{12}^{D_1}\ w_t = (1-B)^r(1-B^{12})\theta_1(B)\Theta_1(B^{12})a_t \quad 7.2.4$$

and we see that the invertibility condition will only be satisfied if $d_1 \geq r$ and $D_1 \geq 1$. However, if w_t is to be stationary then $d_1 \leq r$ and $D_1 \leq 1$. The model (7.2.2) will therefore be stationary and invertible only if n_t has the structure

$$\phi_1(B)\Phi_1(B^{12})\ \nabla^{r}\nabla_{12}\ n_t = \theta_1(B)\Theta_1(B^{12})a_t \quad 7.2.5$$

Equally, the differencing operator $\nabla^{r-1}\nabla_{12}$ could be used to reduce the deterministic components to a constant, C, (see Section 5.2.2) and the resulting process $\tilde{w}_t$ defined by the model

$$\tilde{w}_t = w_t - C = (1-B)^{r-1}(1-B^{12})n_t \quad 7.2.6$$

would be both stationary and invertible only if the errors possess a structure of the form

$$\phi_1(B)\Phi_1(B^{12})\ \nabla^{r-1}\nabla_{12}\ n_t = \theta_1(B)\Theta_1(B^{12})a_t \quad 7.2.7$$

Any error structure based on a lesser degree of differencing than that shown in equation (7.2.5) (for model (7.2.2.)) and in equation (7.2.7) (for model (7.2.6)) will thus produce non-invertibility in the models (7.2.2) and (7.2.6) respectively. For example, the process

$$w_t = (1-B)(1-B^{12})n_t \qquad 7.2.8$$

will be non-invertible if $n_t = a_t$.

Similar conclusions can be drawn regarding the use of the operator ∇_{12}^r in respect of the series X_t represented by the model

$$X_t = m_t\, s_t + n_t \qquad 7.2.9$$

For example, if m_t consists of a linear trend then the model

$$w_t = \nabla_{12}^2 X_t = (1-B^{12})^2\, n_t \qquad 7.2.10$$

will be non-invertible if $n_t = a_t$.

7.3 Analysis of a Generated Series

When the Box-Jenkins forecasting procedure is applied to a series which includes a deterministic component we have seen in Section 7.2 that, for particular error structures, the resulting A.R.I.M.A. model will be non-invertible. We now consider a Box-Jenkins analysis of a series generated by the model

$$X_t = \alpha + \beta t + a_t \qquad 7.3.1$$

Such a process is a special case of the model (7.2.1) and since the errors are assumed to be stationary ($n_t = a_t$), differencing of equation (7.3.1) will produce a non-invertible A.R.I.M.A. model.

Although this thesis is primarily concerned with seasonal forecasting, the above process should be adequate for illustrating the use of the Box-Jenkins procedure on series which include deterministic components.

7.3.1 Generation of an "Artificial" Series

An "artificial" series X_t (t=1,2,3,...,60) was obtained using the model

$$X_t = 200 + 10t + a_t \qquad 7.3.2$$

The sequence of independent random variables a_t, a_{t-1}, a_{t-2},...' was generated from a normal distribution with zero mean and variance 400, using tables from Beyer (1968).

The generated data are tabulated below and plotted in Figure 7.1.

Table 7.1 Data generated from the process $X_t = 200 + 10t + a_t$

t	X_t	t	X_t	t	X_t	t	X_t
1	223	16	377	31	493	46	630
2	221	17	360	32	522	47	688
3	259	18	386	33	492	48	661
4	238	19	417	34	499	49	666
5	280	20	431	35	541	50	723
6	245	21	395	36	572	51	712
7	263	22	411	37	579	52	733
8	255	23	459	38	574	53	738
9	254	24	459	39	608	54	748
10	311	25	446	40	618	55	728
11	298	26	458	41	606	56	755
12	355	27	463	42	621	57	772
13	306	28	488	43	634	58	766
14	340	29	487	44	627	59	791
15	359	30	526	45	664	60	810

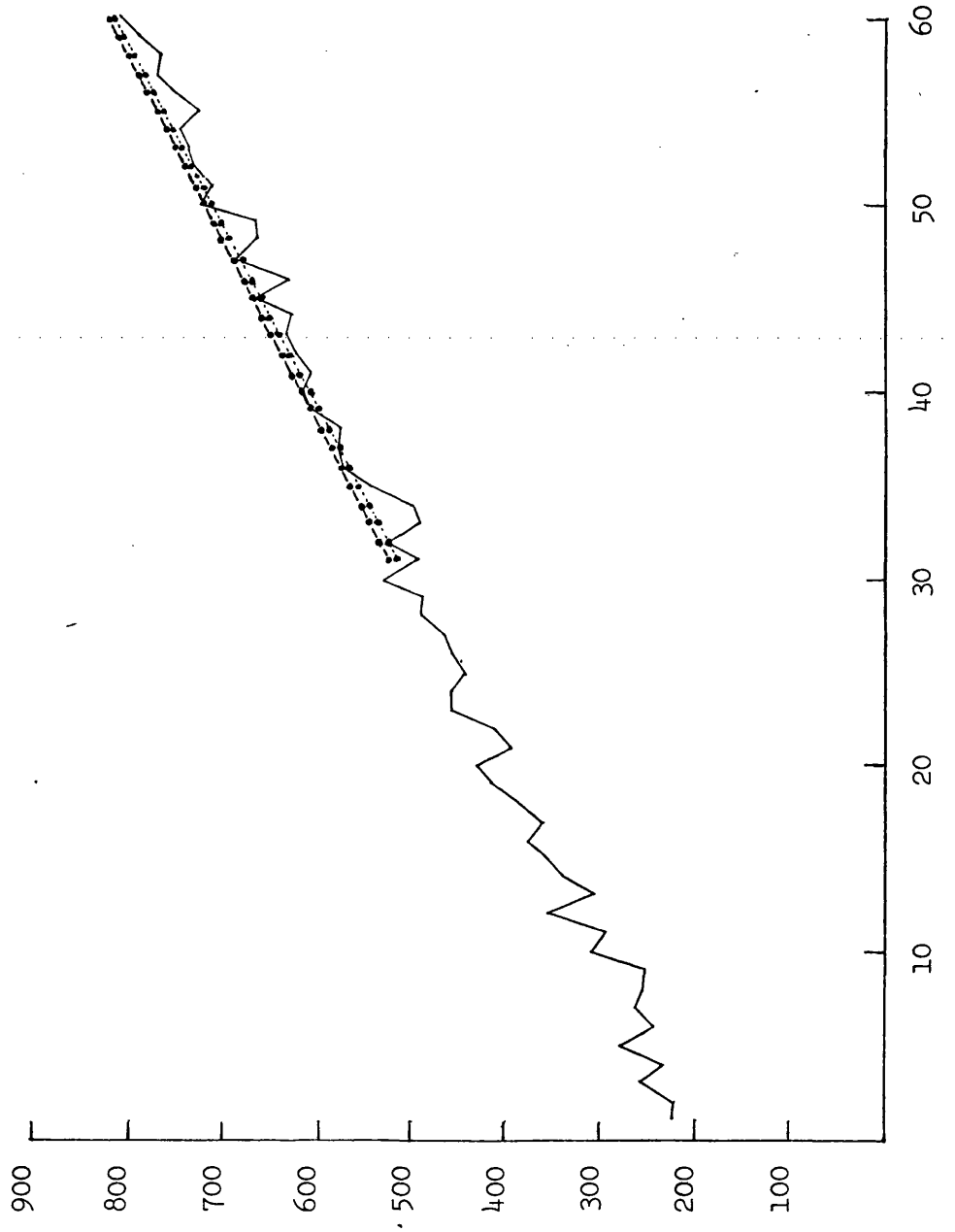


FIGURE 7.1

Generated series together with Box-Jenkins lead time 1 forecasts (•–•–•)
and regression forecasts (•–•–•–•–•)

7.3.2 Box-Jenkins Analysis

Faced with the data generated in Section 7.3.1, a Box-Jenkins analyst would begin by examining the sample autocorrelation function of various differences of the series X_t. Table 7.2 gives the sample autocorrelation functions for X_t and ∇X_t, for lags 1 to 10.

Table 7.2 Sample autocorrelation functions for X_t and ∇X_t

Series	1	2	3	4	5	6	7	8	9	10
X_t	0.94	0.89	0.84	0.80	0.75	0.70	0.65	0.60	0.54	0.49
∇X_t	-0.56	0.15	-0.15	0.07	0.07	-0.24	0.22	-0.09	0.08	-0.10

The ten sample autocorrelation coefficients quoted for the series X_t are all "large" and positive confirming the fact (which is obvious from a visual inspection of the data) that X_t is non-stationary. On the other hand the sample autocorrelation function for ∇X_t dies out quickly, suggesting that ∇X_t is stationary and that no further differencing is necessary.

If the series ∇X_t were completely random, the standard error of the estimated autocorrelations r(k) would be $\frac{1}{\sqrt{59}}$ ($\simeq 0.13$). The absolute value of r(1) is over four times this value and so it can be concluded that the theoretical autocorrelation coefficient at lag 1, $\rho(1)$, is non-zero. Under the assumption that ∇X_t follows a first order moving average process, the standard error of r(k) for $k > 1$ would be approximately $\left(\frac{1}{59}(1 + 2(-0.56)^2)\right)^{\frac{1}{2}} \simeq 0.17$ (see Box and Jenkins (1970, page 34)). Since all the sample auto-correlations for $k > 1$ lie within $\pm 2 \times 0.17$, our tentative choice of model is the first order moving average process. Thus the initially identified model is the A.R.I.M.A. model of order (0,1,1).

Now ∇X_t has a mean which differs significantly from zero so the identified model can be expressed as

$$w_t = \nabla X_t = \theta_o + (1-\theta B)a_t \qquad 7.3.3$$

where Box and Jenkins (1970, page 210) suppose that $\bar{w} = \frac{\sum_{t=1}^{n} w_t}{n}$ is substituted for θ_o, n being the number of observations composing the differenced series w_t. Alternatively θ_o could be estimated simultaneously with θ and this possibly will be examined further in Section 7.4.

Taking $\hat{\theta}_o = \bar{w} = 9.95$, an estimate for θ was obtained by the graphical technique described in Section 2.4.2 and the fitted model was

$$w_t = 9.95 + (1 - 0.97B)a_t \qquad 7.3.4$$

with $\hat{\sigma}_a^2 = 335$.

Despite the fact that $\hat{\theta} = 0.97$, one iteration was sufficient in computing the unconditional sum of squares S(0.97). The reason why further iterative cycles were not necessary can be seen by referring to the equation (6.2.11) used for generating the backward forecast $[w_o]$. The weight given to the term $[e_{n+1}]$ is θ^n which in our case was $(0.97)^{59} \simeq 0.17$ and further the value of $[e_{n+1}]$ computed to start the second iteration was not large.

The diagnostic checks on the residuals, described in Section 2.4.3, were performed and no inadequacies in the fitted model (7.3.4) were indicated.

7.3.3 Comparison with a Regression Analysis

When analysing the same data, Chatfield and Prothero (1973b) say that "the traditional statistician would simply fit a straight

line by least squares and extrapolate". Applying this approach to the whole data, the fitted model was found to be

$$E[X_t] = 203.43 + 9.91t \qquad 7.3.5$$

and the residuals $\hat{e}_t$ were computed using

$$\hat{e}_t = X_t - E[X_t].$$

A residual analysis did not suggest any violation of the assumption that the residuals were independent.

The forecast $\hat{X}_t(\ell)$ made at time t for a period ℓ steps into the future can be generated using the equation

$$\hat{X}_t(\ell) = \underset{t}{E}[X_{t+\ell}] = 203.43 + 9.91 \times (t+\ell)$$

$$= \hat{X}_t(1) + 9.91 \times (\ell-1) \qquad 7.3.6$$

Comparing the fit of the regression model (7.3.5) with that of the A.R.I.M.A. model (7.3.4), the estimated variance of the residuals ($\hat{\sigma}_e^2$) in the regression case was 330 while the Box-Jenkins analysis of Section 7.3.2 yielded an estimated error variance of 335. Thus the model (7.3.5) fitted the data marginally better than the model (7.3.4).

To assess the forecasting performance of each method, the data were divided into two parts. The first 30 observations (probably the minimum requirement for the Box-Jenkins procedure) were used to estimate the parameters in each model and the remainder to assess the relative forecasting potential of the two models.

The re-fitted A.R.I.M.A. model was

$$w_t = \nabla X_t = 10.45 + (1 - 0.99B)a_t \qquad 7.3.7$$

while the equivalent linear regression model was estimated to be

$$E[X_t] = 200.75 + 10.21t \qquad 7.3.8$$

On this occasion at least 4 iterations were needed in computing the unconditional sum of squares S(0.99) during the estimation of the moving average parameter in the model (7.3.7). Obviously the small number of observations employed in the fitting procedure contributed to this occurrence (see Section 7.3.2 and Section 6.2.)

Starting when t = 30, one step ahead forecasts, $\hat{X}_t(1)$, were computed (based on all the observations available at time t) over the next 30 time periods using the fitted models (7.3.7) and (7.3.8). These forecasts are plotted in Figure 7.1. It should be pointed out that in neither case where the parameter estimates updated as more recent observations became available.

Using the linear regression model (7.3.8) the one step ahead forecast made at time t is given by

$$\begin{aligned}\hat{X}_t(1) &= 200.75 + 10.21 \times (t+1) \\ &= \hat{X}_{30}(1) + 10.21 \times (t-30) \qquad 7.3.9\end{aligned}$$

for t=30, 31, 32,...,59, so that in fact these forecasts depend only on the first 30 observations which constitute the fitting period.

On the other hand, the one step ahead forecasts generated by the model (7.3.7) are given by

$$\hat{X}_t(1) = X_t + 10.45 - 0.99\hat{a}_t$$

$$= X_t + 10.45 - 0.99\,(X_t - \hat{X}_{t-1}(1))$$

$$= 0.01\,X_t + 10.45 + 0.99\,\hat{X}_{t-1}(1)$$

$$= 0.01\,(X_t + 0.99\,X_{t-1} + (0.99)^2\,X_{t-2} + \ldots + (0.99)^{t-31}X_{31})$$

$$+ 10.45\,(1 + 0.99 + (0.99)^2 + \ldots + (0.99)^{t-31})$$

$$+ (0.99)^{t-30}\,\hat{X}_{30}(1) \qquad 7.3.10$$

for $t = 31,32,33,\ldots,59$, and so these forecasts depend heavily on the forecast $\hat{X}_{30}(1)$ although rather less so than the forecasts generated by equation (7.3.9).

It can be seen from Figure 7.1 that the regression model forecasts are always less than the Box-Jenkins forecasts. The forecast errors resulting from both models include a bias, the mean of the forecast errors from the model (7.3.7) being -17.6 and that for the model (7.3.8) errors is -12.9. The corresponding mean squared errors for the two methods are

BOX-JENKINS ANALYSIS	583
LINEAR REGRESSION ANALYSIS	439

and so for the data generated in Section 7.3.1 the regression approach performed rather better than the Box-Jenkins procedure.

Thus although in theory the linear regression technique is equivalent to the A.R.I.M.A. model of order (0,1,1) with the moving average parameter unity, in practice the results from the two procedures are not identical, The reasons for this are two-fold. Firstly, adopting the Box-Jenkins procedure it is impossible to obtain an estimate of unity for the moving average parameter since the sum of squares S(1.0) diverges as more iterations are performed. Secondly and more important to the case in question,

the estimate of the parameter θ_o in the A.R.I.M.A. model differed from the estimate of the linear coefficient in the regression analysis. The effect of this difference is examined in Section 7.4.

7.4 A Further Characteristic of the Box-Jenkins Procedure

The analysis described in Section 7.3 drew attention to a further characteristic of the Box-Jenkins procedure when applied to series which include deterministic components.

Let us consider again a series X_t which is a realisation of the process (7.3.1). Employing a linear regression technique, the model fitted to the series X_t would be

$$E[X_t] = \hat{\alpha} + \hat{\beta}t \qquad 7.4.1$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the least squares estimates of the parameters α and β.

Differencing equation (7.3.1), the equivalent A.R.I.M.A. model would be

$$w_t = \nabla X_t = \theta_o + (1 - \theta B)a_t \qquad 7.4.2$$

where $\theta_o = \beta$ and $\theta = 1$. In practice, a value of θ slightly less than unity would be employed. The problems involved in estimating θ have already been dealt with in Chapter 6 and Section 7.3. We now turn our attention to the estimation of θ_o (or β).

In Section 7.3.2 the sample mean of the w_t's, $\bar{w}$, was substituted for β, following the statement of Box and Jenkins (1970, page 210) that "for the sample sizes normally considered in time series analysis, this approximation will be adequate". We will denote this estimator for β by β^*, so that

$$\beta^* = \bar{w} = \frac{\sum_{t=1}^{n} w_t}{n} = \frac{X_N - X_1}{N - 1} \qquad 7.4.3$$

where N and n are the lengths of the series X_t and w_t respectively, i.e. $n = N - 1$. Let us now look at some of the properties of the estimator β^*.

Bias

$$E[\beta^*] = \frac{1}{N-1} E[X_N - X_1]$$

$$= \frac{1}{N-1} E[(\alpha + \beta N + a_N) - (\alpha + \beta + a_1)]$$

$$= \beta$$

Thus β^* is an unbiased estimator for β.

Variance of β^*

$$V[\beta^*] = E[\beta^* - E[\beta^*]]^2$$

$$= E\left[\frac{1}{(N-1)^2} [X_N - X_1]^2\right] - \beta^2$$

$$= \frac{2}{(N-1)^2} \sigma_a^2 \qquad \text{for } N > 1$$

Relative Efficiency of β^* with Respect to the Least Squares Estimator $\hat{\beta}$.

The least squares estimator $\hat{\beta}$ for β used in the model (7.4.1) is given by

$$\hat{\beta} = \frac{N \sum_{t=1}^{N} X_t t - \left[\sum_{t=1}^{N} X_t\right]\left[\sum_{t=1}^{N} t\right]}{N \sum_{t=1}^{N} t^2 - \left[\sum_{t=1}^{N} t\right]^2}$$

$$= \frac{6\{2 \sum_{t=1}^{N} X_t t - (N+1) \sum_{t=1}^{N} X_t\}}{(N-1)\, N\, (N+1)}$$

This estimator for β is of course unbiased and it can be shown that the variance of $\hat{\beta}$ is

$$V[\hat{\beta}] = \frac{12}{(N-1)\, N\, (N+1)}\, \sigma_a^2 \qquad \text{for } N > 1$$

Defining the relative efficiency of β^* with respect to $\hat{\beta}$ by

$$\text{Relative Efficiency (R.E)} = \frac{\text{Variance of } \hat{\beta}}{\text{Variance of } \beta^*}$$

we get

$$\text{R.E.} = \frac{12}{(N-1)N(N+1)}\, \sigma_a^2 \Big/ \frac{2}{(N-1)^2}\, \sigma_a^2$$

$$= \frac{6(N-1)}{N(N+1)} \qquad \text{for } N \geq 2.$$

Values of this quantity for various values of N are given in Table 7.3.

Table 7.3 Relative efficiencies of β^* with respect to $\hat{\beta}$.

N	2	3	4	5	6	7	8	9	10
R.E	1.00	1.00	0.90	0.80	0.71	0.64	0.58	0.53	0.49
N	20	30	40	50	60	70	80	90	100
R.E	0.27	0.19	0.14	0.12	0.10	0.08	0.07	0.07	0.06

From the above table it can be seen that for series long enough to apply the Box-Jenkins procedure, the estimator β^* is relatively much less efficient than $\hat{\beta}$. Even for a series of 30 observations (short by most standards) the relative efficiency is as low as 20%.

These results suggest that in the Box-Jenkins analysis described in Section 7.3.2 and 7.3.3 it would have been better to have estimated the parameter θ_o simultaneously with the θ parameter, i.e. by least squares. This possibility is now explored.

The A.R.I.M.A. model (7.3.3) was re-fitted to the first 30 observations of the data generated in Section 7.3.1, the parameter θ_o being estimated simultaneously with θ. The resulting fitted model was

$$w_t = \nabla X_t = 10.21 + (1 - 0.99B)a_t \qquad 7.4.4$$

and this model was used to generate forecasts over the remaining 30 observations. The mean squared error was computed to be 405 which is less than the mean squared error (439) obtained from the linear regression model (7.3.8) and considerably less than the mean squared error (583) from the fitted A.R.I.M.A. model (7.3.7). Thus when applying the Box-Jenkins procedure to series which include deterministic components it is advisable to estimate the constant term θ_o in the A.R.I.M.A. model by least squares.

We now briefly consider two extensions to the linear model (7.3.1). Firstly we will look at the case when the series X_t includes a quadratic trend and secondly at the case when an additive seasonal component is superimposed on a linear trend.

Quadratic Trend

The following process is considered:

$$X_t = \alpha + \beta t + \gamma t^2 + a_t \qquad 7.4.5$$

The equivalent A.R.I.M.A. model is derived by differencing the equation (7.4.5) twice to give

$$w_t = \nabla^2 X_t = \theta_o + (1-\theta B)^2 a_t \qquad 7.4.6$$

with $\theta_o = 2\gamma$ and $\theta = 1$.

Substituting $\bar{w} = \frac{\sum_{t=1}^{n} w_t}{n}$ for 2γ, the estimator γ^* for γ will be

$$\gamma^* = \frac{\bar{w}}{2} = \frac{X_N - X_{N-1} - X_2 + X_1}{2(N-2)}$$

and the relative efficiency of γ^* with respect to the least squares estimator $\hat{\gamma}$ is

$$R.E = \frac{180(N-2)}{(N-1)N(N+1)(N+2)} \qquad \text{for } N \geq 4$$

Table 7.4 shows values of the relative efficiency for several values of N.

Table 7.4 Relative efficiencies of γ^* with respect to $\hat{\gamma}$

N	4	5	6	7	8	9	10
R.E.	1.00	0.64	0.43	0.30	0.21	0.16	0.12
N	20	30	40	50	60	70	80
R.E.	0.02	0.01	<0.01	<0.01	<0.01	<0.01	<0.01

The relative efficiency of γ^* with respect to $\hat{\gamma}$ decreases rapidly as the length of the series increases. Thus it would appear that the need to use the least squares estimator is even greater than in the linear case encountered earlier.

A Seasonal Model

Finally, let us suppose that the series X_t is a realisation of the seasonal process

$$X_t = \alpha + \beta t + s_t + a_t \qquad 7.4.7$$

where $s_t = s_{t-12}$.

Differencing equation (7.4.7) with respect to the seasonal period leads to the A.R.I.M.A. model

$$w_t = \nabla_{12} X_t = \theta_o + (1-\Theta B^{12})a_t \qquad 7.4.8$$

with $\theta_o = 12\beta$ and $\Theta = 1$.

If $\bar{w}$ is substituted for θ_o, i.e. 12β, then the estimator β^* of β is

$$\beta^* = \frac{\bar{w}}{12} = \frac{(X_N+X_{N-1}+\ldots+X_{N-11}) - (X_{12}+X_{11}+\ldots+X_1)}{12(N-12)}$$

and the relative efficiency of β^* with respect to the least squares estimator $\hat{\beta}$ is given by

$$\text{R.E.} = \frac{72(N-12)}{N(N+12)}$$

$$= \frac{6(m-1)}{m(m+1)} \qquad \text{for } m \geq 2$$

where N is the number of observations in the series X_t and it is assumed that the seasonal cycle is repeated m times, i.e. $N = 12m$.

The above expression is similar to that derived for the non-seasonal process with a deterministic linear trend. The only difference is that the total number of observations N in the non-seasonal case is replaced by the number of times the seasonal cycle is repeated, m. As an example, for a non-seasonal series with 60 observations the relative efficiency of β^* with respect to $\hat{\beta}$ is 0.10 while for monthly data with the same number of observations the relative efficiency is 0.80. To achieve a value of 0.10 in the latter case, 60 years data would have to be available. Hence we conclude that the substitution of $\bar{w}$ for β compares more favourably with the least squares estimate in the case of seasonal data than for non-seasonal data.

If the seasonal cycle has a general period s, then the expression for the relative efficiency of β^* with respect to $\hat{\beta}$ still takes the form

$$\text{R.E.} = \frac{6(m-1)}{m(m+1)} \qquad \text{for } m \geq 2$$

where $N = ms$.

7.5 Conclusions

We began this chapter by considering the sort of series which when differenced, produce non-invertible A.R.I.M.A. models. If a series which includes deterministic trend and seasonal components is reduced to stationarity by the use of a differencing operator which removes these deterministic components, then the resulting A.R.I.M.A. model will be invertible only if the error structure is based on the same degree of differencing. The possible estimation difficulties when a process includes parameters close to the non-invertibility region

were discussed in Chapter 6.

The performance of the Box-Jenkins procedure was compared with that of a linear regression on a generated series which included a deterministic linear trend. Initially the regression technique produced the smaller mean squared error but by employing a different estimator for the constant term θ_o in the A.R.I.M.A. model, the Box-Jenkins procedure did much better and in fact performed slightly better than the regression approach. In Section 7.4 we showed that substituting the mean of the differenced series ($\bar{w}$) for θ_o (as in the original Box-Jenkins analysis) is relatively much less efficient than estimating θ_o by least squares. Similar results were obtained for a series which included a quadratic trend and for a seasonal process. Clearly it is advisable to use the least squares estimator for θ_o when analysing series which include deterministic components.

Although the Box-Jenkins procedure eventually generated rather more accurate forecasts than the linear regression technique, the difference between the two mean squared errors was not large. Further, while the estimation of the parameters in the A.R.I.M.A. model necessitated a numerical least squares procedure, the regression model parameters were estimated analytically. Thus in situations where linear or quadratic deterministic trends are suspected, the use of a traditional regression model would be preferred to the equivalent A.R.I.M.A. model. On the other hand, when it is only possible to describe a time series <u>locally</u> by a trend-seasonal model it would be inappropriate to use the regression approach as employed in Section 7.3.3. Instead, the more flexible Box-Jenkins procedure, or a method such as that proposed by Winters (1960) in which the trend and seasonal factors are updated, should be used.

CHAPTER 8

A.R.I.M.A. MODELS ARISING FROM AGGREGATES OF SEVERAL STOCHASTIC PROCESSES

8.1 Introduction

The additive model (7.2.1), for describing a time series X_t, has been employed previously in Sections 5.2.2 and 7.2. On each occasion both the trend and seasonal components were assumed to be deterministic. In Section 5.2.2 we discussed the differencing operators capable of removing the deterministic components from the series X_t and in Section 7.2 we examined the error structures for which the resulting A.R.I.M.A. model would be non-invertible.

Box and Jenkins (1970, page 92) say that "the assumption of a stochastic trend is often more realistic than the assumption of a deterministic trend" and so in this chapter we will consider a model similar to model (7.2.1) but with the trend component being assumed to be stochastic. It will also be assumed that the seasonal component possesses a stochastic structure so that the series X_t can be represented algebraically as

$$X_t = m_t + s_t + n_t \qquad 8.1.1$$

where m_t and s_t are stochastic trend and seasonal components and n_t is an extraneous error term. The series X_t can thus be regarded as the aggregate of three independent stochastic processes.

Our attention will be focussed on the relationship between the degree of differencing required to reduce X_t to stationarity and the differencing operators on which the trend, seasonal (where applicable) and error components are based. In addition, we shall also see how an A.R.I.M.A. process can result from the aggregate of several less

complicated A.R.I.M.A. processes, a possibility explored by Box and Jenkins (1970, page 121) and Granger (1972).

In Section 8.2 we will examine the A.R.I.M.A. model arising when the non-seasonal version of the model (8.1.1) is assumed while a special case will be looked at in detail in Section 8.3. The seasonal case will be investigated in Section 8.4 and Section 8.5 will contain a summary of the results obtained from the preceeding sections.

8.2 Non-seasonal Processes

Let us consider a particular case of the model (8.1.1) viz. the model

$$X_t = m_t + n_t \qquad 8.2.1$$

It is assumed that the trend is described by the A.R.I.M.A. model

$$\phi_1(B)\ \nabla^{d_1}\ m_t = \theta_1(B)\ a_t' \qquad 8.2.2$$

where $\phi_1(B)$ and $\theta_1(B)$ are polynomials of order p_1 and q_1 respectively and a_t' is a white noise process with variance $\sigma^2_{a'}$.

The extraneous errors n_t are considered to be represented by the A.R.I.M.A. model

$$\phi_2(B)\ \nabla^{d_2}\ n_t = \theta_2(B)\ a_t'' \qquad 8.2.3$$

$\phi_2(B)$ and $\theta_2(B)$ being of order p_2 and q_2 and a_t'' is a white noise process (variance $\sigma^2_{a''}$) mutually independent of a_t'.

Now if d is the order of differencing necessary to reduce the series X_t to a stationary series w_t then from equation (8.2.1)

$$w_t = \nabla^d X_t = \nabla^d m_t + \nabla^d n_t \qquad 8.2.4$$

and substituting from equations (8.2.2) and (8.2.3) leads to

$$\phi_1(B)\ \phi_2(B)\ (1-B)^{d_1+d_2} w_t = (1-B)^{d+d_2}\ \phi_2(B)\ \theta_1(B) a_t'$$

$$+ (1-B)^{d+d_1}\ \phi_1(B)\ \theta_2(B) a_t'' \qquad 8.2.5$$

If $d_1 \geq d_2$, then equation (8.2.5) can be written as

$$\phi_1(B)\ \phi_2(B)\ (1-B)^{d_1+d_2} w_t = (1-B)^{d+d_2} \{\phi_2(B)\ \theta_1(B) a_t'$$

$$+ (1-B)^{d_1-d_2}\ \phi_1(B)\ \theta_2(B) a_t''\} \qquad 8.2.6$$

Since it has been assumed that w_t is stationary, then

$$d_1 + d_2 \leq d + d_2 \quad \text{i.e.} \quad d \geq d_1$$

and also for the process (8.2.6) to be invertible

$$d_1 + d_2 \geq d + d_2 \quad \text{i.e.} \quad d \leq d_1$$

Hence for the model (8.2.6) to satisfy both the stationarity and invertibility conditions we must have $d = d_1$.

Similarly, if $d_2 \geq d_1$ we arrive at $d = d_2$ and so in general the degree of differencing necessary to reduce X_t to stationarity and at the same time give rise to an invertible A.R.M.A. model, is given by

$$d = \max(d_1, d_2)$$

When such a degree of differencing is employed, Granger (1972) shows that the stationary process w_t is described by the A.R.M.A. model of order (p,q) where

$$p \leq p_1 + p_2$$

and

$$q \leq \max(d_1 - d_2 + p_1 + q_2,\ p_2 + q_1) \text{ if } d_1 \geq d_2$$

$$\leq \max(d_2 - d_1 + p_2 + q_1,\ p_1 + q_2) \text{ if } d_2 \geq d_1$$

The need for the inequalities in the above expressions arises partly from the fact that $\phi_1(B)$ and $\phi_2(B)$ may contain common roots (see Granger (1972)).

8.3 The A.R.I.M.A. Model of Order (0,1,1)

Let us suppose that m_t and n_t are described by the models

$$\left.\begin{aligned} \nabla m_t &= a_t' \\ n_t &= a_t'' \end{aligned}\right\} \qquad 8.3.1$$

This formulation has been adopted by Muth (1960) and it assumes that the observed time series X_t is made up of two components, one lasting a single time period (a_t'') and the other through all subsequent periods (m_t). The former is referred to as the transitory component, the latter as the permanent component.

Using the result obtained in Section 8.2, the series $w_t = \nabla X_t$ can be represented by a model which is both stationary and invertible. From equation (8.2.6)

$$w_t = \nabla X_t = a_t' + (1-B)a_t'' \qquad 8.3.2$$

and the autocovariance function for the w_t's is

$$\gamma_w(k) = \begin{cases} \sigma_{a'}^2 + 2\sigma_{a''}^2 & k = 0 \\ -\sigma_{a''}^2 & k = 1 \\ 0 & k \geq 2 \end{cases}$$

The only non-zero autocovariance (apart from $\gamma_w(0)$) is at lag 1 so that w_t can be described by the first order moving average process

$$w_t = (1 - \theta B)a_t \qquad 8.3.3$$

The parameters θ and σ_a^2 associated with the model (8.3.4) are

related to σ_a^2, and σ_a^2,, by

$$(1 + \theta^2)\ \sigma_a^2 = \sigma_a^2, + 2\ \sigma_a^2,,$$

$$-\theta\sigma_a^2 = -\sigma_a^2,, \qquad 8.3.4$$

which on solving yields

$$\theta = \frac{(\sigma_a^2'/\sigma_a^2,, + 2) \pm \sqrt{\sigma_a^4'/\sigma_a^4,, + 4\ \sigma_a^2'/\sigma_a^2,,}}{2}$$

whence $\sigma_a^2 = \dfrac{\sigma_a^2,,}{\theta}$

Generally two values for θ satisfy the equations (8.3.4) but only the smaller will be inside the unit circle. Thus θ must take the value

$$\theta = \frac{(\sigma_a^2'/\sigma_a^2,, + 2) - \sqrt{\sigma_a^4'/\sigma_a^4,, + 4\ \sigma_a^2'/\sigma_a^2,,}}{2}$$

which will always lie in the region $0 \leq \theta \leq 1$. When $\theta = 1$ the model (8.3.3) is of course non-invertible and such a situation will arise when σ_a^2, = 0 (see equation (8.3.4)). This implies that the trend component is deterministic and given by m_t = a constant.

On the other hand, if m_t = a constant, σ_a^2, = 0 and using equation (8.3.4) we get

$$(1 - \theta)^2\ \sigma_a^2 = 0$$

but $\sigma_a^2 \neq 0$ and so $\theta = 1$. Thus a necessary and sufficient condition for the differencing operation to produce a model which is non-invertible is that the trend should simply be a constant.

The other extreme value of θ is zero and from equation (8.3.4) this occurs when σ_a^2,, = 0 i.e. when the distribution of n_t is concentrated entirely on the point $n_t = 0$.

In Section (9.4.2) we shall see that the forecast generated by the model (8.3.3) is the same as the simple exponential smoothing forecast

so that the expression for θ quoted above is equivalent to that for the optimal smoothing constant derived by Muth (1960). In practice $\sigma^2_{a'}$ and $\sigma^2_{a''}$ will not generally be known and θ would be estimated by the means described in Section 2.4 and Chapter 6. The forms of m_t and n_t defined by equation (8.3.1) do however deomonstrate one way in which the A.R.I.M.A. model of order (0,1,1) can arise. Alternative structures for m_t and n_t which also lead to the model (8.3.3) are

$$\left.\begin{aligned} \nabla m_t &= a'_t - \theta_1 a'_{t-1} \\ n_t &= a''_t \end{aligned}\right\} \qquad 8.3.5$$

and

$$\left.\begin{aligned} \nabla m_t &= a'_t - \theta_1 a'_{t-1} \\ \nabla n_t &= a''_t - \theta^*_1 a''_{t-1} \end{aligned}\right\} \qquad 8.3.6$$

In the case of the formulation (8.3.6), when $\theta_1 = \theta^*_1 = 0$ X_t would be described by the A.R.I.M.A. model of order (0,1,0).

Generally, extending the approach followed in this section, if

$$\left.\begin{aligned} \nabla^{d_1} m_t &= a'_t \\ n_t &= a''_t \end{aligned}\right\} \qquad 8.3.7$$

then the aggregate series X_t can be represented by the A.R.I.M.A. model of order $(0,d_1,d_1)$.

8.4 Seasonal Processes

The results derived in Section 8.2 are now extended to the case in which the seasonal model (8.1.1) is considered. As in Section 8.2 it will be assumed that the trend and error terms are described by the A.R.I.M.A. models (8.2.2) and (8.2.3) respectively. Additionally we suppose that the seasonal component s_t follows the general A.R.I.M.A. process of order $(P_1,D_1,Q_1)_{12}$

$$\Phi_1(B^{12})\ \nabla_{12}^{D_1}\ s_t = \theta_1(B^{12})\ a_t''' \qquad 8.4.1$$

where a_t''' is a white noise process (variance $\sigma^2_{a'''}$) mutually independent of a_t' and a_t''.

If the differencing operator $\nabla^d\ \nabla_{12}^D$ is required to reduce the aggregate series X_t to a stationary series w_t, then

$$w_t = \nabla^d\ \nabla_{12}^D\ X_t = \nabla^d\ \nabla_{12}^D\ m_t + \nabla^d\ \nabla_{12}^D\ s_t + \nabla^d\ \nabla_{12}^D\ n_t \qquad 8.4.2$$

Substituting for m_t, s_t and n_t from equations (8.2.2), (8.4.1) and (8.2.3) leads to

$$\begin{aligned}
&\phi_1(B)\ \phi_2(B)\ \Phi_1(B^{12})\ (1-B)^{d_1+\ d_2}\ (1-B^{12})^{D_1}\ w_t \\
&\quad = (1-B)^d\ (1-B^{12})^D\ \{\phi_2(B)\ \Phi_1(B^{12})\ (1-B)^{d_2}\ (1-B^{12})^{D_1}\ \theta_1(B)\ a_t' \\
&\quad + \phi_1(B)\ \phi_2(B)\ (1-B)^{d_1+\ d_2}\ \theta_1(B^{12})\ a_t''' \\
&\quad + \phi_1(B)\ \Phi_1(B^{12})\ (1-B)^{d_1}\ (1-B^{12})^{D_1}\ \theta_2(B)\ a_t''\} \qquad 8.4.3
\end{aligned}$$

If $d_2 + D_1 \leq d_1 + d_2,\ d_1 + D_1$, then from equation (8.4.3) it can be seen that for the process w_t to be stationary and invertible d and D must satisfy

$$d = d_1 - D_1\ ,\ D = D_1$$

Similarly, if $d_1 + d_2 \leq d_2 + D_1,\ d_1 + D_1$, then

$$d = 0\ \ ,\ \ D = D_1$$

while if $d_1 + D_1 \leq d_2 + D_1,\ d_1 + d_2$, we have

$$d = d_2 - D_1\ \ ,\ \ D = D_1$$

Thus the differencing operators necessary for the process defined by equation (8.4.3) to be both stationary and invertible must be of orders

$$d = \max(d_1 - D_1,\ 0,\ d_2 - D_1)\ ,\ D = D_1.$$

The autoregressive operator in the A.R.M.A. model for describing w_t is of order p where

$$p \leq p_1 + p_2 + 12\,P_1$$

but the order of the moving average operator, q, depends very much on the relative values of d_1, d_2 and D_1. For example, when $d_2 + D_1 \leq d_1 + d_2$, $d_1 + D_1$

$$\begin{aligned} q \leq \max(&p_2 + q_1 + 12\,P_1 + 11\,D_1, \\ &p_1 + p_2 + d_1 + 12\,Q_1 - D_1, \\ &p_1 + q_2 + d_1 - d_2 + 12\,P_1 + 11\,D_1) \end{aligned}$$

We illustrate the above results by considering the structures

$$\left.\begin{aligned} \nabla^2 m_t &= a_t' \\ \nabla_{12} s_t &= a_t''' \\ n_t &= a_t'' \end{aligned}\right\} \qquad 8.4.4$$

d and D must be chosen so that $d = \max(1,0,-1) = 1$ and $D = 1$, hence

$$\begin{aligned} w_t = \nabla\nabla_{12}X_t &= (1+B+B^2+\ldots.+B^{11})a_t' + (1-B)a_t''' + (1-B)(1-B^{12})a_t'' \\ &= (1-\theta_1 B-\theta_2 B^2 \ldots\ldots - \theta_{13}B^{13})a_t \end{aligned} \qquad 8.4.5$$

The aggregate process X_t is therefore described by a multi-parameter integrated moving average model which is not in the general class of multiplicative seasonal models. In practice the fitting of an A.R.I.M.A. model with 13 moving average parameters would probably never be attempted even if one strongly believed a time series to be generated by the models (8.1.1) and (8.4.4). Instead we would try to explain the non-zero autocorrelations at lags 1 to 13 by a low order autoregressive, moving average or mixed model. For example, if $\sigma^2_{a,,}$ is much larger than both $\sigma^2_{a,}$ and $\sigma^2_{a,,,}$ then the A.R.I.M.A. model of order $(0,1,1) \times (0,1,1)_{12}$

might provide an adequate approximation. Table 8.1 shows the autocorrelation function for the model (8.4.5) when $\sigma^2_{a,,} = 20$, $\sigma^2_{a'} = \sigma^2_{a,,,} = 1$ and also the autocorrelation function for the model

$$w_t = (1 - 0.36B)\ (1 - 0.57B^{12})a_t \qquad 8.4.6$$

with $\sigma^2_a = 63$.

The autocorrelation function for the model (8.4.6) agrees reasonably closely with that of the model (8.4.5) since the autocorrelations in the latter case are all rather small between lags 2 and 10. In practice it would be very difficult to decide whether a sample autocorrelation function could be associated with the model (8.4.5) or (8.4.6), in which case the simpler model would be assumed.

8.5 Summary

If a time series which includes a deterministic component is reduced to stationarity by a differencing operator which removes the deterministic part, then we saw in Section 7.2 that for certain error structures the resulting A.R.M.A. model may be non-invertible. The situation in which a series is assumed to be composed of several independent stochastic (as opposed to deterministic) components is rather different. We found that for stochastic trend and seasonal components it is always possible to select a differencing operator which will produce a series describable by a stationary, invertible A.R.M.A. model. Thus the estimation difficulties discussed in Chapter 6 will occur less frequently if the series in question is generated by several independent stochastic processes than if deterministic components are present.

In addition to examining the choice of differencing operator necessary to produce a stationary, invertible process, this chapter also demonstrated how a quite complicated A.R.I.M.A. process can be generated from the sum of several simpler processes. In particular, for non-seasonal models, if the d_1^{th} difference of the trend component is a

Table 8.1 Autocorrelation functions for the models (8.4.5) and (8.4.6)

model \ lag	1	2	3	4	5	6	7	8	9	10	11	12	13
(8.4.5)	-0.32	0.11	0.10	0.09	0.07	0.06	0.05	0.04	0.03	0.02	0.22	-0.43	0.21
(8.4.6)	-0.32	0	0	0	0	0	0	0	0	0	0.14	-0.43	0.14

white noise process and the additional error component another independent white noise process then the aggregate series X_t can be represented by the A.R.I.M.A. model of order $(0,d_1,d_1)$.

Our examination of seasonal processes revealed that in theory the aggregate series was generally described not by a model in the multiplicative seasonal class but by a multi-parameter A.R.I.M.A. model. However, in practice it may prove difficult to decide whether a particular sample autocorrelation function is associated with a complex model or a much simpler model. In such a case the latter model would be identified.

CHAPTER 9

INTERPRETATION OF A.R.I.M.A. MODELS

9.1 Introduction

As we have seen in preceeding chapters, the Box-Jenkins forecasting procedure essentially involves the fitting of a particular A.R.I.M.A. model which adequately describes the series to be forecasted. On the other hand, the forecasting techniques proposed by such authors as Winters (1960), Brown (1963) and Harrison (1965) assume that the series in question can be described locally by a more traditional trend-seasonal model such as that shown in equation (7.2.1). The resulting forecasts can then be expressed in terms of exponentially weighted moving averages (E.W.M.A.'s).

Generally, in the Box-Jenkins case, it is not immediately obvious what the A.R.I.M.A. model is telling us about the data in terms of trend and seasonality and hence the interpretation of the derived forecasts is not clear. This point has been commented upon by Chatfield and Prothero (1973a). Box and Jenkins (1973) take the view that A.R.I.M.A. models "are usually rather easy to understand" and illustrate their point by expressing the forecast generated by one particular A.R.I.M.A. model as an E.W.M.A.

The major part of this chapter consists of an attempt at representing the general form of the A.R.I.M.A. model in terms of E.W.M.A.'s . Expressions for individual models can then be obtained by substitution in the general form.

It is of course a matter of personal opinion whether A.R.I.M.A. models are any easier to understand when expressed as E.W.M.A.'s and other representations may be preferred. One other alternative approach is considered in this chapter: namely that of expressing the A.R.I.M.A. model directly in terms of trend and seasonality.

Section 9.2 contains a brief discussion of E.W.M.A.'s and also of how they have been adapted for forecasting purposes by Brown (1963) The interpretation of A.R.I.M.A. models in terms of E.W.M.A.'s will be dealt with in Sections 9.3 (stationary processes), 9.4 (non-stationary processes) and 9.5 (seasonal processes). At each stage special cases will be examined. In Section 9.6 we shall see how the interpretation is affected by the fact that, in practice, series are not infinite in extent.

The interpretation of A.R.I.M.A. models will be considered from rather a different angle in Section 9.7. It will be shown that by solving the difference equation which constitutes the A.R.I.M.A. model, a representation in terms of trend and seasonal models can be achieved.

The conclusions arrived at from the results obtained in this chapter will be stated in Section 9.8.

9.2 Exponentially Weighted Moving Averages

The first order E.W.M.A. at time t, $\bar{z}_t^{(A_1)}$, of a time series $z_t, z_{t-1}, z_{t-2}, \ldots$ is defined to be

$$\bar{z}_t^{(A_1)} = A_1 \sum_{j=o}^{\infty} (1-A_1)^j z_{t-j} \qquad 9.2.1$$

where the super script (A_1) indicates that the data have been smoothed once with respect to the smoothing constant A_1 $(0 < A_1 < 1)$.

Smoothing the series $\bar{z}_t^{(A_1)}, \bar{z}_{t-1}^{(A_1)}, \bar{z}_{t-2}^{(A_1)}, \ldots$ with respect to a second smoothing constant A_2, we obtain the second order E.W.M.A. $\bar{z}_t^{(A_1,A_2)}$, i.e.

$$\bar{z}_t^{(A_1,A_2)} = A_2 \sum_{k=o}^{\infty} (1-A_2)^k \bar{z}_{t-k}^{(A_1)}$$

$$= A_1 A_2 \sum_{j=o}^{\infty} \sum_{k=o}^{\infty} (1-A_1)^j (1-A_2)^k z_{t-j-k} \qquad 9.2.2$$

Generally, the m^{th} order E.W.M.A. $\underset{\cdot}{\underset{-}{z}}_t(A_1,A_2,\ldots,A_m)$ can be defined by

$$\underset{\cdot}{\underset{-}{z}}_t(A_1,A_2,\ldots,A_m) = A_1A_2\ldots A_m \sum_{i_1=o}^{\infty}\sum_{i_2=o}^{\infty}\ldots\sum_{i_m=o}^{\infty}(1-A_1)^{i_1}(1-A_2)^{i_2}\ldots (1-A_m)^{i_m} z_{t-i_1-i_2-\ldots-i_m} \qquad 9.2.3$$

In his discussion of exponential smoothing, Brown (1963) assumes that the same smoothing constant is employed for all orders of smoothing i.e. $A_1 = A_2 = \ldots = A_m\ (= A)$.

Using the method outlined in Section 1.2.2, Brown (1963) shows that when a series z_t can be described by a polynomial function, the resulting forecasts can be expressed in terms of the first m^{th} order E.W.M.A.'s. The value of m depends on the degree of the polynomial. If z_t can be represented locally by the constant model

$$z_t = \alpha_o(t) + e_t \qquad 9.2.4$$

where the coefficient $\alpha_o(t)$ is to be estimated at time t and e_t is a random error, then the optimal one step ahead forecast $\hat{z}_t(1)$ is simply

$$\hat{z}_t(1) = \underset{-}{z}_t(A)$$

Extending model (9.2.4) to the local linear representation

$$z_t = \alpha_o(t) + \alpha_1(t)t + e_t \qquad 9.2.5$$

Brown (1963) shows that the one step ahead forecast is given by

$$\hat{z}_t(1) = \left[\frac{2-A}{1-A}\right]\underset{-}{z}_t(A) - \frac{1}{(1-A)}\underset{-}{z}_t(A,A)$$

while for the local quadratic model

$$z_t = \alpha_o(t) + \alpha_1(t)t + \frac{\alpha_2(t)}{2} t^2 + e_t \qquad 9.2.6$$

the one step ahead forecast is

$$\hat{z}_t(1) = \left[\frac{A^2 - 3A + 3}{(1-A)^2}\right] \bar{z}_t(A) - \left[\frac{3-A}{(1-A)^2}\right] \bar{\bar{z}}_t(A,A) + \frac{1}{(1-A)^2} \bar{\bar{\bar{z}}}_t(A,A,A)$$

More generally, Brown (1963, page 133) goes on to prove that if the local model is

$$z_t = \sum_{i=o}^{n} \alpha_i(t)t^i + e_t \qquad 9.2.7$$

then the one step ahead forecast can be expressed as a linear combination of the first (n+1) orders of E.W.M.A.'s.

A comprehensive account of general exponential smoothing is given by Brown (1963) while most of the formulae of practical importance are quoted by Kendall (1973).

9.3 Stationary Models

This section deals with the interpretation (in terms of E.W.M.A.'s, where possible) of autoregressive models, moving average models and mixed autoregressive-moving average models. In each case the series z_t is assumed to have a zero mean.

9.3.1 Autoregressive Models

If z_t is a series described by the autoregressive model of order p, defined by equation (2.2.2), then the one step ahead forecast made at time t, $\hat{z}_t(1)$, is given by

$$\hat{z}_t(1) = \phi_1 z_t + \phi_2 z_{t-1} + \ldots + \phi_p z_{t-p+1} \qquad 9.3.1$$

This forecast is therefore a linear combination of the p most recent observations and does not involve E.W.M.A.'s. When p = 1

$$\hat{z}_t(1) = \phi_1 z_t \quad , \qquad |\phi_1| < 1$$

and the forecast is simply a fraction of the value of the most recent observation.

9.3.2 Moving Average Models

The interpretation of moving average models is rather more difficult than that of autoregressive models. In the former the emphasis is on past values of the white noise process while for the latter the models are expressed in terms of actual observations. In order to examine the interpretation of the moving average model of order q (see equation (2.2.3)) it is best to write this model as

$$z_t = (1-H_1B)(1-H_2B) \ldots (1-H_qB)a_t \qquad 9.3.2$$

where $|H_j| < 1$, for j = 1,2,3,...,q, to ensure invertibility.

Theoretically, the H_j's may be real or any pair of roots H_i H_j may be complex. For reasons which will become apparent later we shall confine ourselves to the case when all the H_j's (j = 1,2,3,...,q) are real.

Multiplying both sides of equation (9.3.2) by $(1-H_qB)^{-1}$ we get

$$\begin{aligned} z_t &= - H_q z_{t-1} - H_q^2 z_{t-2} - \ldots + (1-H_1B)(1-H_2B)\ldots(1-H_{q-1}B)a_t \\ &= - (1-A_q)z_{t-1} - (1-A_q)^2 z_{t-2} - \ldots + (1-H_1B)(1-H_2B)\ldots(1-H_{q-1}B)a_t \end{aligned} \qquad 9.3.3$$

where $A_q = 1 - H_q$.

Now if $0 < H_q < 1$ then $0 < A_q < 1$ and we can write

$$\underline{z}_t^{(A_q)} = A_q \sum_{j=o}^{\infty} (1-A_q)^j z_{t-j}$$

Thus using equation (9.3.3)

$$z_t = -\frac{H_q}{A_q} \underline{z}_{t-1}^{(A_q)} + (1-H_1B)(1-H_2B) \ldots (1-H_{q-1}B) a_t \qquad 9.3.4$$

Such a rperesentation is not possible for negative or complex values of H_q since $\underline{z}_{t-1}^{(A_q)}$ is not defined for these values.

Both sides of equation (9.3.4) can be multiplied by $(1-H_{q-1}B)^{-1}$ to give, provided $0 < H_{q-1} < 1$,

$$z_t = -\frac{H_{q-1}}{A_{q-1}} \underline{z}_{t-1}^{(A_{q-1})} - \frac{H_q}{A_{q-1}A_q} \underline{\underline{z}}_{t-1}^{(A_{q-1},A_q)} + (1-H_1B)(1-H_2B) \ldots (1-H_{q-2}B) a_t \qquad 9.3.5$$

with $A_{q-1} = 1 - H_{q-1}$.

Thus, if $0 < H_j < 1$, for $j = 1,2,3,\ldots,q$, then continuing the above operation a further q-2 times the following expression for the one step ahead forecast made at time t is arrived at:

$$\hat{z}_t(1) = -\frac{H_1}{A_1} \underline{z}_t^{(A_1)} - \frac{H_2}{A_1A_2} \underline{\underline{z}}_t^{(A_1,A_2)} \ldots - \frac{H_q}{A_1A_2..A_q} \underline{\underline{\vdots}}_t^{(A_1,A_2,\ldots,A_q)} \qquad 9.3.6$$

where $A_j = 1 - H_j$ $(j = 1,2,3,\ldots,q)$.

The one step ahead forecast generated by a q^{th} order moving average model is therefore a linear combination of the most recent first, second, ..., qth order E.W.M.A.'s.

As an example, the one step ahead forecast for a first order moving average process is

$$\hat{z}_t(1) = \frac{-\theta_1}{1-\theta_1} \underline{z}_t^{(A_1)}$$

i.e. a multiple of the most recent first order E.W.M.A. with smoothing constant $A_1 = 1-\theta_1$. This representation is only possible for θ_1 in the range $0 < \theta_1 < 1$ when the corresponding range for $-\theta_1/1-\theta_1$ is shown in Figure 9.1.

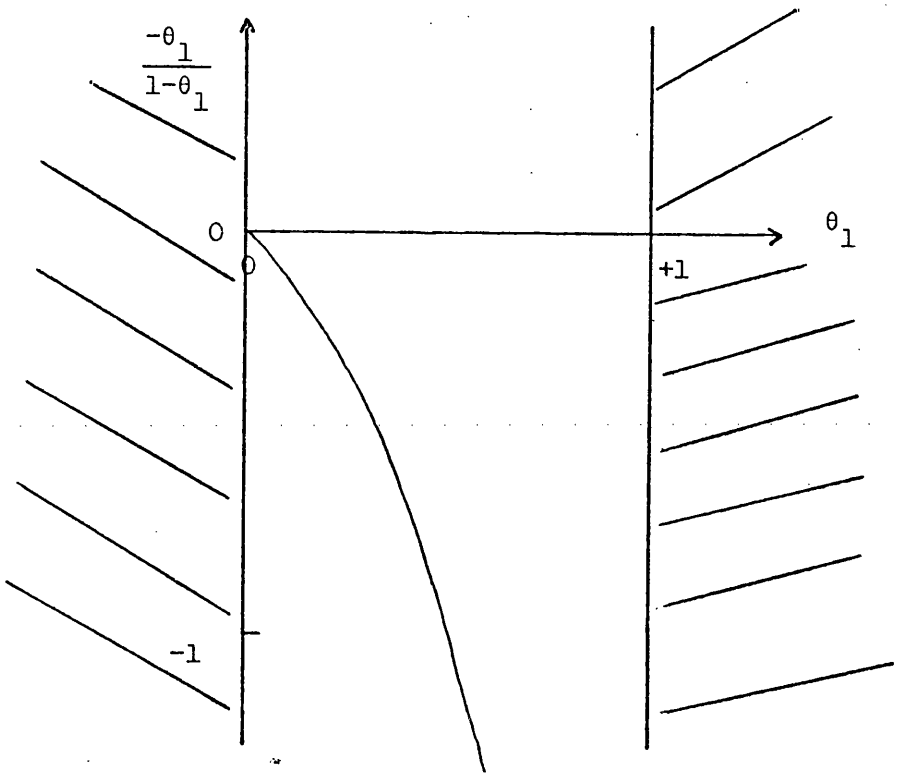


Fig. 9.1. Range of values taken by $-\theta_1/(1-\theta_1)$

As θ_1 increases in the region 0 to $\frac{1}{2}$, $\frac{-\theta_1}{1-\theta_1}$ decreases to -1 and then as θ_1 approaches unity $\frac{-\theta_1}{1-\theta_1}$ tends to $-\infty$. This may seem rather surprising but it should be remembered that as $\theta_1 \to 1$ so $\bar{z}_t^{(A_1)} \to 0$.

9.3.3 Mixed Autoregressive - Moving Average Models

As for the moving average models dealt with in the previous section, it is more convenient to express the autoregressive-moving average model of order (p,q), usually defined by equation (2.2.5), in the form

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + (1-H_1B)(1-H_2B)\ldots(1-H_qB)a_t$$

Multiplying both sides of equation (9.3.7) successively by $(1-H_qB)^{-1}$, $(1-H_{q-1}B)^{-1}$,...,$(1-H_1B)^{-1}$, as in Section 9.3.2, then provided all the H_j's are real and satisfy $0 < H_j < 1$, the one step ahead forecast $\hat{z}_t(1)$ can be expressed in the form

$$\hat{z}_t(1) = -\frac{H_1}{A_1}\underline{z}_t^{(A_1)} - \frac{H_2}{A_1A_2}\underline{z}_t^{(A_1,A_2)} - \ldots - \frac{H_q}{A_1A_2\ldots A_q}\underline{\dot{z}}_t(A_1,A_2,\ldots,A_q)$$

$$+ \frac{1}{A_1A_2\ldots A_q}\left[\phi_1\,\underline{\dot{z}}_t(A_1,A_2,\ldots,A_q) + \phi_2\,\underline{\dot{z}}_{t-1}(A_1,A_2,\ldots,A_q)\right.$$

$$\left. + \ldots + \phi_p\,\underline{\dot{z}}_{t-p+1}(A_1,A_2,\ldots,A_q)\right] \qquad 9.3.8$$

where again $A_j = 1 - H_j$ ($j = 1,2,3,\ldots,q$).

Thus $\hat{z}_t(1)$ is a linear combination of the first q^{th} order E.W.M.A.'s at time t and the q^{th} order E.W.M.A.'s for the p most recent time periods.

Alternatively, by making use of the relationship

$$\underline{\dot{z}}_t(A_1,A_2,\ldots,A_k) = A_k\,\underline{\dot{z}}_t(A_1,A_2,\ldots,A_{k-1}) + (1-A_k)\,\underline{\dot{z}}_{t-1}(A_1,A_2,\ldots,A_k)$$

$$\text{for } k = 1,2,3,\ldots,q$$

recursively, a different form of the equation (9.3.8) can be obtained. If $p > q$ then $\hat{z}_t(1)$ can be expressed as a linear combination of the first q^{th} order E.W.M.A.'s at time t plus the (p-q) most recent observations. If $q \geq p$, the one step ahead forecast is a linear combination of just the first q^{th} order E.W.M.A.'s at time t.

For the A.R.M.A. model of order (1,1), equation (9.3.8) reduces to

$$\hat{z}_t(1) = \frac{-\theta_1}{1-\theta_1}\underline{z}_t^{(A_1)} + \frac{1}{1-\theta_1}\phi_1\,\underline{z}_t^{(A_1)}$$

$$= \left(\frac{\phi_1-\theta_1}{1-\theta_1}\right)\underline{z}_t^{(A_1)} \qquad (A_1 = 1-\theta_1)$$

As for the first order moving average process, the one step ahead forecast is some multiple of the most recent first order E.W.M.A., although on this occasion the weight given to $\underline{z}_t^{(A_1)}$ can be positive or negative, depending on the value of $\phi_1-\theta_1$.

9.4 Non-stationary Models

The expressions derived in Section 9.3 for stationary models will now be extended to cover non-stationary processes. The general A.R.I.M.A. model of order (p,d,q) will be dealt with first and then several important special cases will be considered.

9.4.1 The General A.R.I.M.A. Model

The general A.R.I.M.A. model of order (p,d,q) has been defined by equation (2.3.2). Once again the moving average operator is re-parameterised in terms of its zeros and it is also advantageous to express the stationary autoregressive operator and the differencing operator in terms of the non-stationary operator $\phi^*_{p+d}(B)$. Thus

$$(1-\phi^*_1B-\phi^*_2B^2 - \ldots - \phi^*_{p+d}B^{p+d})z_t = (1-H_1B)(1-H_2B)\ldots(1-H_qB)\,a_t \qquad 9.4.1$$

and using the notation adopted in equation (9.3.8) the forecast $\hat{z}_t(1)$ can (for real H_j and $0 < H_j < 1$ $(j = 1,2,3,\ldots,q)$) be expressed as

$$\begin{aligned}\hat{z}_t(1) = &-\frac{H_1}{A_1}\underline{z}_t^{(A_1)} - \frac{H_2}{A_1A_2}\underline{z}_t^{(A_1,A_2)} - \ldots - \frac{H_q}{A_1A_2\ldots A_q}\underline{\dot{z}}_t^{(A_1,A_2,\ldots,A_q)} \\ &+ \frac{1}{A_1A_2\ldots A_q}\Big[\phi^*_1\underline{\dot{z}}_t^{(A_1,A_2,\ldots,A_q)} + \phi^*_2\underline{\dot{z}}_t^{(A_1,A_2,\ldots,A_q)} \\ &+ \ldots + \phi^*_{p+d}\underline{\dot{z}}_{t-p-d+1}^{(A_1,A_2,\ldots,A_q)}\Big] \qquad 9.4.2\end{aligned}$$

The lead time one forecast made at time t is therefore a linear combination of the first q^{th} order E.W.M.A.'s at time t and the q^{th} order E.W.M.A.'s for the p+d most recent time periods. Employing the argument used in Section 9.3.3, $\hat{z}_t(1)$ has an alternative representation. If $p + d > q$, then the forecast can be expressed as a linear combination of the first q^{th} order E.W.M.A.'s at time t and the (p+d-q) most recent observations. On the other hand if $p + d \leq q$ then $\hat{z}_t(1)$ can be expressed as a linear combination of the first q^{th} order E.W.M.A.'s at time t, only.

9.4.2 The A.R.I.M.A. Model of Order(0,1,1)

If $0 < \theta_1 < 1$, substitution in the general expression (9.4.2) results in

$$\hat{z}_t(1) = \frac{-\theta_1}{1-\theta_1} \underline{z}_t^{(A_1)} + \frac{1}{1-\theta_1} \underline{z}_t^{(A_1)} = \underline{z}_t^{(A_1)} \qquad 9.4.3$$

and the one step ahead forecast is simply the most recent first order E.W.M.A.. This forecast will be identical to that obtained by Brown (1963), when the constant model (9.2.4) is appropriate, provided the smoothing constant $A = A_1 = 1 - \theta_1$. Box and Jenkins (1970, pages 106-108) also refer to this result.

9.4.3 The A.R.I.M.A. Model of Order (0,2,2)

Provided $0 < H_j < 1$ $(j = 1,2)$, we see from equation (9.4.2) that the one step ahead forecast generated by the A.R.I.M.A. model of order (0,2,2) is

$$\hat{z}_t(1) = -\frac{H_1}{A_1} \underline{z}_t^{(A_1)} - \frac{H_2}{A_1 A_2} \underline{z}_t^{(A_1,A_2)} + \frac{1}{A_1 A_2} \left[2 \underline{z}_t^{(A_1,A_2)} - \underline{z}_{t-1}^{(A_1,A_2)} \right] \qquad 9.4.4$$

However

$$\underline{\underline{z}}_t^{(A_1,A_2)} = A_2 \underline{z}_t^{(A_1)} + (1-A_2) \underline{\underline{z}}_{t-1}^{(A_1,A_2)}$$

and so $$\underline{\underline{z}}_{t-1}^{(A_1,A_2)} = \frac{1}{(1-A_2)} \left[\underline{\underline{z}}_t^{(A_1,A_2)} - A_2 \underline{z}_t^{(A_1)} \right]$$

Substituting for $\underline{\underline{z}}_{t-1}^{(A_1,A_2)}$ in equation (9.4.4) and setting $H_j = 1 - A_j$ $(j = 1,2)$ results in

$$\hat{z}_t(1) = -\frac{(1-A_1)}{A_1} \underline{z}_t^{(A_1)} - \frac{(1-A_2)}{A_1A_2} \underline{\underline{z}}_t^{(A_1,A_2)}$$

$$+ \frac{1}{A_1A_2} \left[2 \underline{\underline{z}}_t^{(A_1,A_2)} - \frac{1}{(1-A_2)} \left(\underline{\underline{z}}_t^{(A_1,A_2)} - A_2 \underline{z}_t^{(A_1)} \right) \right]$$

$$= \left[\frac{1 - (1-A_1)(1-A_2)}{A_1(1-A_2)} \right] \underline{z}_t^{(A_1)} - \left[\frac{A_2}{A_1(1-A_2)} \right] \underline{\underline{z}}_t^{(A_1,A_2)} \qquad 9.4.5$$

which is a linear combination of the first and second order E.W.M.A.'s at time t. Now if $A_1 = A_2 = A$, where A is the smoothing constant associated with Brown's method, then

$$\hat{z}_t(1) = \left[\frac{2-A}{1-A} \right] \underline{z}_t^{(A)} - \frac{1}{(1-A)} \underline{\underline{z}}_t^{(A,A)} \qquad 9.4.6$$

This is precisely the forecast quoted in Section 9.2 obtained by exponential smoothing, when the linear model (9.2.5) is assumed. Thus for a linear model, exponential smoothing produces forecasts which are identical to those generated by a special case of the A.R.I.M.A. model of order (0,2,2) viz. the model in which $H_1 = H_2 = 1 - A$ or $\theta_1 = 2(1-A)$ and $\theta_2 = (1-A)^2$.

Similarly, it follows that in the case of the quadratic model (9.2.6), the exponential smoothing forecast given in Section 9.2 is the same as that generated by the A.R.I.M.A. model of order (0,3,3)

in the particular instance when $H_1 = H_2 = H_3 = 1-A$ or $\theta_1 = 3(1-A)$, $\theta_2 = -3(1-A)^2$ and $\theta_3 = (1-A)^3$. Again, A is the smoothing constant associated with Brown's method.

Box and Jenkins (1970, pages 168-170) give an example where Brown (1963) assumed the quadratic model (9.2.6) whereas the appropriate A.R.I.M.A. model was of order (0,1,1). The latter model generated considerably better forecasts than those obtained using Brown's method, simply because the initial choice of model by Brown was a wrong one. However, even if a quadratic model had been appropriate, the A.R.I.M.A. model of order (0,3,3) would have yielded a mean squared forecast error theoretically at least as small as that produced by exponential smoothing.

So far we have concentrated on the interpretation of A.R.I.M.A. models which include no autoregressive parameters. Let us now look at a simple case in which an autoregressive parameter is present.

9.4.4 The A.R.I.M.A. Model of Order (1,1,1)

For the A.R.I.M.A. model of order (1,1,1), substitution in the general expression (9.4.2) gives

$$\hat{z}_t(1) = \frac{-\theta_1}{1-\theta_1} \underline{z}_t^{(A_1)} + \frac{1}{(1-\theta_1)} [(1+\phi_1)\underline{z}_t^{(A_1)} - \phi_1 \underline{z}_{t-1}^{(A_1)}]$$

$$= \left[\frac{1+\phi_1-\theta_1}{1-\theta_1}\right] \underline{z}_t^{(A_1)} - \frac{\phi_1}{(1-\theta_1)} \underline{z}_{t-1}^{(A_1)} \qquad 9.4.7$$

i.e. the one step ahead forecast is a weighted average of the two most recent first order E.W.M.A.'s. Alternatively, making use of the relationship

$$\underline{z}_t^{(A_1)} = A_1 z_t + (1-A_1) \underline{z}_{t-1}^{(A_1)}$$

$$= (1-\theta_1) z_t + \theta_1 \underline{z}_{t-1}^{(A_1)}$$

$\hat{z}_t(1)$ can be expressed as

$$\hat{z}_t(1) = \left[1 - \frac{\phi_1}{\theta_1} \right] \bar{z}_t^{(A_1)} + \frac{\phi_1}{\theta_1} z_t \qquad 9.4.8$$

which is a weighted average of the most recent first order E.W.M.A. and the most recent observation. If $\theta_1 < 2\phi_1$ then more weight is attached to z_t than to $\bar{z}_t^{(A_1)}$ while if $\theta_1 > 2\phi_1$ the reverse is true.

Employing Brown's method in the case when the A.R.I.M.A. model of order (1,1,1) is appropriate, the constant model (9.2.4) would be assumed and the resulting forecast $\hat{z}_t^B(1) = \bar{z}_t^{(A)}$. Thus the Box-Jenkins forecast (9.4.8) is related to Brown's forecast by

$$\hat{z}_t(1) = \hat{z}_t^B(1) + \left[1 - \frac{\phi_1}{\theta_1} \right] \bar{z}_t^{(A_1)} + \frac{\phi_1}{\theta_1} z_t - \hat{z}_t^B(1)$$

and if $A = A_1 = 1-\theta_1$

$$\hat{z}_t(1) = \hat{z}_t^B(1) + \frac{\phi_1}{\theta_1} \left[z_t - \bar{z}_t^{(A_1)} \right]$$

$$= \hat{z}_t^B(1) + \phi_1 \left[z_t - \hat{z}_{t-1}^B(1) \right] \qquad 9.4.9$$

The forecast generated by the model of order (1,1,1) would therefore be Brown's forecast adjusted by a fraction of the error in the previous Brown forecast.

9.5 Seasonal Models

In this section we will derive an expression, in terms of E.W.M.A.'s., for the general multiplicative seasonal model (2.6.5), although for convenience a seasonal period of 12 will be assumed. The interpretation of the forecasts derived from the four models A, B, C and D encountered in Section 3.7 will be dealt with in detail.

9.5.1 The General Multiplicative Seasonal Model of Order $(p,d,q) \times (P,D,Q)_{12}$

The multiplicative seasonal model defined by equation (2.6.4) can be re-stated as

$$\begin{aligned}(1-\Phi_1^*B^{12} - \Phi_2^*B^{24} - \ldots - \Phi_{P+D}^*B^{12(P+D)})\, z_t \\ = (1 - \Phi_1 B^{12} - \Phi_2 B^{24} - \ldots - \Phi_P B^{12P})\ (1-B^{12})^D\, z_t \\ = (1 - J_1B^{12})(1-J_2B^{12}) \ldots (1-J_QB^{12})\, \alpha_t \end{aligned} \qquad 9.5.1$$

where

$$(1 - \theta_1B^{12} - \theta_2B^{24} - \ldots - \theta_QB^{12Q}) = \prod_{i=1}^{Q} (1-J_iB^{12})$$

and α_t follows the process

$$\begin{aligned}(1 - \phi_1^*B - \phi_2^*B^2 - \ldots - \phi_{p+d}^*B^{p+d})\, \alpha_t \\ = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)(1-B)^d\, \alpha_t \\ = (1 - H_1B)(1-H_2B) \ldots (1-H_qB)\, a_t \end{aligned} \qquad 9.5.2$$

Using equation (9.4.2), remembering that the α_t's are not generally uncorrelated, the one step ahead forecast generated by the model (9.5.1) is

$$\begin{aligned}\hat{z}_t(1) = &- \frac{J_1}{C_1}\, \bar{z}_{t-11}(C_1) - \frac{J_2}{C_1C_2}\, \bar{z}_{t-11}(C_1,C_2) - \ldots - \frac{J_Q}{C_1C_2..C_Q}\, \bar{z}_{t-11}(C_1,C_2,\ldots C_Q) \\ &+ \frac{1}{C_1C_2..C_Q} \Big[\Phi_1^*\, \dot{\bar{z}}_{t-11}(C_1,C_2,\ldots C_Q) + \Phi_2^*\, \dot{\bar{z}}_{t-23}(C_1,C_2,\ldots,C_Q) \\ &\qquad + \ldots + \Phi_{P+D}^*\, \dot{\bar{z}}_{t-12(P+D)+1}(C_1,C_2,\ldots,C_Q) \Big] \\ &+ E_t\left[\alpha_{t+1}\right] \end{aligned} \qquad 9.5.3$$

where for example

$$\underline{z}_{t-11}^{(C_1)} = C_1 \sum_{i=o}^{\infty} (1-C_1)^i \, z_{t-11-12i}$$

and $C_j = 1 - J_j \quad (j = 1,2,3,\ldots,Q)$.

Again using equation (9.4.2)

$$E[\alpha_{t+1}] = \hat{\alpha}_t(1) = -\frac{H_1}{A_1}\underline{\alpha}_t^{(A_1)} - \frac{H_2}{A_1A_2}\underline{\alpha}_t^{(A_1,A_2)} - \ldots - \frac{H_q}{A_1A_2\ldots A_q}\underline{\dot{\alpha}}_t^{(A_1,A_2\ldots,A_q)}$$

$$+ \frac{1}{A_1A_2\ldots A_q}\left[\phi_1^*\underline{\dot{\alpha}}_t^{(A_1,A_2\ldots,A_q)} + \phi_2^*\underline{\dot{\alpha}}_{t-1}^{(A_1,A_2,\ldots,A_q)}\right.$$

$$\left. + \ldots + \phi_{p+d}^*\underline{\dot{\alpha}}_{t-p-d+1}^{(A_1,A_2,\ldots,A_q)}\right] \qquad 9.5.4$$

with

$$\underline{\dot{\alpha}}_{t-j}^{(A_1,A_2,\ldots,A_k)} = \frac{1}{C_1C_2\ldots C_Q}\left[\underline{\dot{z}}_{t-j}^{(A_1,A_2,\ldots,A_k)(C_1,C_2,\ldots,C_Q)}\right.$$

$$\left. - \phi_1^*\underline{\dot{z}}_{t-12-j}^{(A_1,A_2,\ldots,A_k)(C_1,C_2,\ldots,C_Q)} - \ldots - \phi_{P+D}^*\underline{\dot{z}}_{t-12(P+D)-j}^{(A_1,,A_2,\ldots,A_k)(C_1,C_2,\ldots,C_Q)}\right]$$

9.5.5

for $k = 1,2,3,\ldots,q$, $j = 0,1,2,\ldots,p+d-1$.

The notation $\underline{\dot{z}}_{t-j}^{(A_1,A_2,\ldots,A_K)(C_1,C_2,\ldots,C_Q)}$ refers to the k^{th} order E.W.M.A. smoothed a further Q times with respect to the seasonal period.

For example

$$\underline{z}_{t-j}^{(A_1)(C_1)} = C_1 \sum_{i_1=o}^{\infty} (1-C_1)^{i_1} \underline{z}_{t-12i_1-j}^{(A_1)}$$

$$= C_1A_1 \sum_{i_1=o}^{\infty}\sum_{i_2=o}^{\infty} (1-C_1)^{i_1}(1-A_1)^{i_2} z_{t-12i_1-i_2-j}$$

The expressions (9.5.3), (9.5.4) and (9.5.5) which define the lead time l forecast are obviously too involved to interpret in the general case. However these expressions are useful in the sense that the forecast generated by a particular model can be expressed in terms of E.W.M.A.'s by substituting for p,d,q,P,D and Q in the general form. This point is illustrated in the following subsections.

9.5.2 The A.R.I.M.A. Model of Order $(1,1,0) \times (0,1,1)_{12}$

The interpretation of the A.R.I.M.A. model of order $(1,1,0)\times(0,1,1)_{12}$ has been discussed by Box and Jenkins (1973), for particular values of ϕ_1 and θ_1. We now consider the understanding of this model for any values of ϕ_1 and θ_1.

Substitution in equations (9.5.3),(9.5.4) and (9.5.5) leads to

$$\begin{aligned}\hat{z}_t(l) &= \bar{z}^{(C_1)}_{t-1l} + \frac{(1+\phi_1)}{C_1}\left[\bar{z}^{(C_1)}_{t} - \bar{z}^{(C_1)}_{t-12}\right] \\ &\quad - \frac{\phi_1}{C_1}\left[\bar{z}^{(C_1)}_{t-1} - \bar{z}^{(C_1)}_{t-13}\right] \\ &= \bar{z}^{(C_1)}_{t-1l} + (1+\phi_1)\left[z_t - \bar{z}^{(C_1)}_{t-12}\right] \\ &\quad - \phi_1\left[z_{t-1} - \bar{z}^{(C_1)}_{t-13}\right] \qquad 9.5.6\end{aligned}$$

where $C_1 = 1 - \theta_1$.

The forecast for, say, June made in May would therefore be the first order yearly E.W.M.A. for the previous June adjusted by a weighted average of the differences between the most recent May figure and the previous May's yearly E.W.M.A. and the most recent April figure and the previous April's yearly E.W.M.A.

Thus if an alternative forecasting technique based on the simple forecast

$$\hat{z}^{s}_t(l) = \bar{z}^{(C)}_{t-1l} = C \sum_{i=o}^{\infty} (1-C)^i \, z_{t-1l-12i}$$

is employed then, provided $C = C_1 = 1 - \theta_1$, the forecast generated by the A.R.I.M.A. model of order $(1,1,0) \times (0,1,1)_{12}$ can be regarded as the simple forecast adjusted by a weighted average of the two most recent forecast errors resulting from the alternative technique.

9.5.3 The A.R.I.M.A. Model of Order $(1,1,0) \times (1,1,0)_{12}$

For the A.R.I.M.A. model of order $(1,1,0) \times (1,1,0)_{12}$

$$\hat{z}_t(1) = (1+\phi_1)\, z_t - \phi_1\, z_{t-1} + (1+\Phi_1)\, z_{t-11} - (1+\phi_1)(1+\Phi_1)\, z_{t-12} + \phi_1(1+\Phi_1)\, z_{t-13} - \Phi_1\, z_{t-23} + \Phi_1(1+\phi)\, z_{t-24} - \phi_1\Phi_1\, z_{t-25} \qquad 9.5.7$$

and so, for example, the forecast made in May for June is a linear combination of the most recent May and April figures and the June, May and April figures for the previous two years. This forecast can be interpreted in two ways. If we consider the simple forecast

$$\hat{z}^s_t(1) = (1+\phi_1)\, z_t - \phi_1\, z_{t-1} \qquad 9.5.8$$

$\hat{z}_t(1)$ can be written as

$$\hat{z}_t(1) = \hat{z}^s_t(1) + (1+\Phi_1)\,[z_{t-11} - \hat{z}^s_{t-12}(1)] - \Phi_1\,[z_{t-23} - \hat{z}^s_{t-24}(1)]$$

and the forecast generated by the A.R.I.M.A. model of order $(1,1,0) \times (1,1,0)_{12}$ is the forecast defined by equation (9.5.8) adjusted by a weighted average of the errors, resulting from the use of the simpler forecast, for the corresponding periods one and two years previously.

On the other hand in terms of the forecast

$$\hat{z}^s_t(1) = (1+\Phi_1)\, z_{t-11} - \Phi_1 z_{t-23} \qquad 9.5.9$$

we have

$$\hat{z}_t(1) = \hat{z}^s_t(1) + (1+\phi_1)\,[z_t - \hat{z}^s_{t-1}(1)] - \phi_1\,[z_{t-1} - \hat{z}^s_{t-2}(1)]$$

and the forecast (9.5.9) is adjusted by a weighted average of the two most recent forecast errors.

9.5.4 The A.R.I.M.A. Model of Order $(0,1,1) \times (1,1,0)_{12}$

The one step ahead forecast generated by the A.R.I.M.A. model of order $(0,1,1) \times (1,1,0)_{12}$ can be expressed as

$$\hat{z}_t(1) = \bar{z}^{(A_1)}_t + (1+\Phi_1)\,[z_{t-11} - \bar{z}^{(A_1)}_{t-12}] - \Phi_1\,[z_{t-23} - \bar{z}^{(A_1)}_{t-24}] \qquad 9.5.10$$

with $A_1 = 1 - \theta_1$

Thus the forecast for June made in May is the first order monthly E.W.M.A. for May adjusted by a weighted average of the differences between the June figures and the preceeding monthly May E.W.M.A.'s in the previous two years. Once θ_1 and Φ_1 have been estimated, one step ahead forecasts can therefore be generated using simple exponential smoothing and then adding to this forecast a weighted average of the simple exponential smoothing forecast errors for the relevant period one and two years previously.

9.5.5 The A.R.I.M.A. Model of Order $(0,1,1) \times (0,1,1)_{12}$

In the case of the A.R.I.M.A. model of order $(0,1,1) \times (0,1,1)_{12}$ the one step ahead forecast $\hat{z}_t(1)$ is

$$\hat{z}_t(1) = \bar{z}^{(C_1)}_{t-11} + \bar{z}^{(A_1)}_t - \bar{z}^{(A_1)(C_1)}_{t-12} \qquad 9.5.11$$

$(A_1 = 1-\theta_1,\ C_1 = 1-\Theta_1)$

i.e. the forecast for June from May is the sum of the first order monthly E.W.M.A. for May and the first order yearly E.W.M.A. for the previous June, less the first order monthly, first order yearly E.W.M.A. for the previous May. The interpretation of the term $\underline{z}_{t-12}^{(A_1)(C_1)}$ is not obvious and a clearer understanding can be achieved by re-writing equation (9.5.11) in the form

$$\hat{z}_t(1) = \underline{z}_{t-11}^{(C_1)} + \overline{\left\{z_t - \underline{z}_{t-12}^{(C_1)}\right\}}^{(A_1)}$$

so that the one step ahead forecast is the first order yearly E.W.M.A. adjusted by the simple E.W.M.A. of the errors obtained from employing $\underline{z}_{t-11}^{(C_1)}$ as a forecast. By a similar argument the forecast (9.5.11) can be interpreted in terms of the simple E.W.M.A. $\underline{z}_{t-1}^{(A_1)}$.

Thus it has been possible to relate the one step ahead forecasts generated by various seasonal A.R.I.M.A. models to those based on the more familiar E.W.M.A.'s. Although this improves the understanding of A.R.I.M.A. models, other means of interpretation may well be more illuminating. One alternative will be examined in Section 9.7. However before concluding our discussion of E.W.M.A.'s we will investigate a point of practical importance.

9.6 A Practical Consideration

The expressions derived in Sections 9.3, 9.4 and 9.5 are all based on the assumption that an infinite amount of data is available. In practice of course A.R.I.M.A. models are fitted to series which are composed of a finite number of observations. We now see how the interpretation of the resulting forecasts is affected by the fact that series encountered in practice are not infinite in extent.

In Section 9.4.2 we found that for the A.R.I.M.A. model of order (0,1,1) the one step ahead forecast $\hat{z}_t(1)$ was simply

$$\hat{z}_t(1) = \bar{z}_t^{(A_1)} = (1-\theta_1) \sum_{j=o}^{\infty} \theta_1^j z_{t-j} \qquad 9.6.1$$

where $A_1 = 1-\theta_1$

If however the series in question consists of the N observations z_o, z_1, z_2, ..., z_n (n = N-1), then the forecast generated by the model of order (0,1,1) would be

$$\hat{z}_n(1) = (1-\theta_1) \sum_{j=o}^{n-1} \theta_1^j z_{n-j} + \theta_1^n z_o - \theta_1^{n+1} \hat{a}_o \qquad 9.6.2$$

where $\hat{a}_o$ is the estimate of the initial residual.

Provided θ_1 is not close to $\pm$ 1 (see Chapter 6) then for moderate values of n the terms θ_1^{n+j} (j > 0) will be negligible and so to a good approximation the forecast defined by equation (9.6.2) can be written as

$$\hat{z}_n(1) \simeq (1-\theta_1) \sum_{j=o}^{n-1} \theta_1^j z_{n-j} \simeq (1-\theta_1) \sum_{j=o}^{\infty} \theta_1^j z_{n-j} = \bar{z}_n^{(A_1)}$$

Thus, even allowing for the fact that in practice the data will be finite in extent, the one step ahead forecast generated by the A.R.I.M.A. model of order (0,1,1) can still be generally regarded as an E.W.M.A.

As pointed out by Chatfield and Prothero (1973b), the same conclusion may not be true for seasonal data. Let us consider the seasonal model

$$(1-B^{12})\, z_t = (1-\theta_1 B^{12})\, a_t \qquad 9.6.3$$

Using equation (9.5.3), for $0 < \theta_1 < 1$, the one step ahead forecast is theoretically given by

$$\hat{z}_t(1) = \bar{z}_{t-11}^{(C_1)} = (1-\theta_1) \sum_{j=o}^{\infty} \theta_1^j z_{t-11-12j} \qquad 9.6.4$$

where $C_1 = 1-\theta_1$.

In practice, for a series possessing the N observations $z_{-11}, z_{-10}, z_{-9}, \ldots, z_n$ (n=N-12), in which the seasonal cycle is repeated m times (i.e. N=12m), the one step ahead forecast made at time t = n is

$$\hat{z}_n(1) = (1-\theta_1) \sum_{j=o}^{m-2} \theta_1^j z_{n-11-12j} + \theta_1^{m-1} z_{-11} - \theta_1^m \hat{a}_{-11} \qquad 9.6.5$$

Unlike the non-seasonal case, θ_1^m may not be negligible even for series of moderate length and so it will often be necessary to compute the term $\theta_1^m \hat{a}_{-11}$. If the estimate $\hat{a}_{-11}$ is set equal to zero then equation (9.6.5) becomes

$$\hat{z}_n(1) = (1-\theta_1) \sum_{j=o}^{m-2} \theta_1^j z_{n-11-12j} + \theta_1^{m-1} z_{-11} \qquad 9.6.6$$

and quite considerable weight can be given to the remotest observation z_{-11}. Alternatively, $\hat{a}_{-11}$ can be computed by the technique of back forecasting described in Section 2.4.2 and Chapter 6. Employing just one iterative cycle, the backward forecast $[w_{-11}]$ (where $w_t = z_t - z_{t-12}$) is obtained from equation (6.2.13) as

$$[w_{-11}] = -\theta_1 \sum_{j=o}^{m-2} \theta_1^{m-2-j} w_{n-11-12j}$$

whence, from equation (9.6.3) with $\hat{a}_{-23}=0$,

$$\hat{a}_{-11} = -\theta_1 \sum_{j=o}^{m-2} \theta_1^{m-2-j} w_{n-11-12j}$$

$$= -\theta_1 \sum_{j=o}^{m-2} \theta_1^{m-2-j} \left[z_{n-11-12j} - z_{n-11-12(j+1)} \right]$$

Substituting for $\hat{a}_{-11}$ in equation (9.6.5) leads to

$$\hat{z}_n(1) = \left[1-\theta_1+\theta_1^{2m-1}\right] z_{n-11} + \theta_1(1-\theta_1)\left[1+\theta_1^{2m-3}\right] z_{n-23}$$

$$+ \theta_1^2(1-\theta_1)\left[1+\theta_1^{2m-5}\right] z_{n-35} + \ldots + \theta_1^{m-1}(1-\theta_1^2) z_{-11} \qquad 9.6.7$$

The weights given to z_{n-11}, z_{n-23}, z_{n-35}, ..., z_{-11} when $N = 72$ ($m = 6$) and $\theta_1 = 0.8$ are shown in Table 9.1, for the following cases:

(a) when an infinite amount of past data is assumed (equation (9.6.4)),

(b) when $\hat{z}_n(1)$ is obtained from equation (9.6.6), and

(c) when $\hat{z}_n(1)$ is obtained from equation (9.6.7).

Table 9.1. Weights given to past observations when $N = 72$, $\theta = 0.8$

Case	z_{49}	z_{37}	z_{25}	z_{13}	z_1	z_{-11}
(a)	0.20	0.16	0.13	0.10	0.08	0.07
(b)	0.20	0.16	0.13	0.10	0.08	0.33
(c)	0.29	0.18	0.16	0.13	0.12	0.12

From the above table it can be seen that an E.W.M.A. (case (a)) would assign about 74% of its total weight to the observations z_{49}, z_{37}, z_{25}, ..., z_{-11}. The weight given to unknown observations is therefore by no means negligible. A forecast resulting from equation (9.6.6) (case (b)) would depend heavily on the most remote observation z_{-11} (about 1/3 of the total weight is given to z_{-11}). On the other hand, when back forecasting is used (case (c)), progressively less weight is attached to observations farther in the past although the weights do not decrease geometrically. Hence for seasonal models, under certain conditions (small m, realtively large θ_1), the one step ahead forecast is not, in practice, an E.W.M.A.

9.7 Interpretation in Terms of Classical Time Series Models

Thus far we have considered the interpretation of the forecasts generated by A.R.I.M.A. models, solely in terms of E.W.M.A.'s. In this section we explore another way in which the A.R.I.M.A. model

can be more easily understood.

In Chapter 5 (also Section 7.2) we examined the particular A.R.I.M.A. models to be expected when the Box-Jenkins forecasting procedure was applied to series which included deterministic trend and seasonal components. We now look at the reverse procedure, viz. solving the difference equation which constitutes the A.R.I.M.A. model to obtain a model in terms of trend and seasonal components.

If in the general multiplicative seasonal model of order $(p,d,q) \times (P,D,Q)_{12}$, defined by equation (2.6.4), we introduce the operators $\phi^*_u(B)$ of order $u(= p+12P+d+12D)$ and $\theta^*_v(B)$ of order $v(= q+12Q)$ where

$$\phi^*_u(B) = \phi_p(B)\ \Phi_P(B^{12})(1-B)^d(1-B^{12})^D$$

and

$$\theta^*_v(B) = \theta_q(B)\ \Theta_Q(B^{12})$$

then the model (2.6.4) can be written as

$$\phi^*_u(B)\ z_t = \theta^*_v(B)\ a_t \qquad 9.7.1$$

The general solution of equation (9.7.1) can be considered, in relation to the time when the process was first observed (t=0), as the sum

$$z_t = z'_t + z''_t \qquad 9.7.2$$

where z'_t is a particular integral and z''_t the complementary function. The particular integral represents the component which is unpredictable at time t=0 and Box and Jenkins (1970, page 118) show that for $t > v$ the equation (9.7.1) has a particular integral defined by

$$z'_t = \begin{cases} 0 & \text{for } t \leq 0 \\ a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots + \psi_{t-1} a_1 & \text{for } t > 0 \end{cases} \qquad 9.7.3$$

with the ψ weights satisfying the equation

$$\phi_u^*(B)\ \psi(B)\ a_t = \theta_v^*(B)\ a_t \qquad 9.7.4$$

The complementary function z_t'' represents the component of z_t already determined at time t = 0 and it is evaluated by solving the homogeneous difference equation

$$\phi_u^*(B) = 0 \qquad 9.7.5$$

Now $\phi_u^*(B)$ can be expressed as

$$\phi_u^*(B) = (1-G_1B)(1-G_2B)\ldots(1-G_uB) \qquad 9.7.6$$

and when $G_1, G_2, \ldots, G_u$ are distinct z_t'' takes the form

$$z_t'' = \alpha_1 G_1^t + \alpha_2 G_2^t + \ldots + \alpha_u G_u^t \qquad 9.7.7$$

where the α_i's (i = 1,2,3,...,u) are constants. If however the first k roots of $\phi_u^*(B)$ are equal (= G_1) and the remaining u-k roots are distinct then

$$z_t'' = G_1^t \sum_{j=1}^{k} \alpha_j\, t^{j-1} + \sum_{j=k+1}^{u} \alpha_j\, G_j^t \qquad 9.7.8$$

If more than one of the G_j's are repeated then the equation (9.7.8) can be modified in an obvious way.

In general therefore the complementary function will consist of polynomials and a mixture of damped exponentials and damped sine waves (depending whether a root is real or a pair of roots is complex).

It should be stressed that the general solution of equation (9.7.1) can be related to any time K < t. The observation z_t can then be expressed as the sum of the component which is unpredictable at time t = K (the particular integral) and the component already determined at time t = K (the complementary function). The coefficients α_i (i = 1,2,3,...,u) in the complementary function change for different values of K.

We now look more closely at the particular forms of the model (9.7.1) previously dealt with in Sections 9.5.5, 9.5.2, 9.5.4 and 9.5.3 (in that order). In each case K will be taken as zero.

9.7.1 The A.R.I.M.A. Model of Order $(0,1,1) \times (0,1,1)_{12}$

The A.R.I.M.A. model of order $(0,1,1) \times (0,1,1)_{12}$ has a particular integral (for $t > 13$) given by equation (9.7.3) where the ψ weights satisfy

$$(1-B)(1-B^{12})\ \psi(B)\ a_t = (1-\theta_1 B)(1-\theta_1 B^{12})\ a_t$$

The complementary function z_t'' obtained using equation (9.7.8) is

$$z_t'' = \alpha_1 + \alpha_2 t + \alpha_3 e^{\frac{-i2\pi t}{12}} + \alpha_4\, e^{\frac{-i4\pi t}{12}} + \ldots + \alpha_{13}\, e^{\frac{-i22\pi t}{12}}$$

$$= \alpha_1 + \alpha_2 t + (-1)^t \left\{ \frac{c_o}{2} + \sum_{\nu=1}^{5} \left[c_\nu \cos \frac{2\pi\nu t}{12} + d_\nu \sin \frac{2\pi\nu t}{12} \right] \right\}$$

$$= \alpha_1 + \alpha_2 t + s_t \qquad 9.7.9$$

where $s_t = (-1)^t \left\{ \frac{c_o}{2} + \sum_{\nu=1}^{5} \left[c_\nu \cos \frac{2\pi\nu t}{12} + d_\nu \sin \frac{2\pi\nu t}{12} \right] \right\}$

and $s_t = s_{t-12}$, $\sum_{t=1}^{12} s_t = 0$. The term s_t thus represents the seasonal variation in the series z_t and so the complementary function (9.7.9) consists of a linear trend and an additive seasonal component.

It is of interest to compare the complementary function of the above model with those associated with the models examined in Sections 9.5.2 to 9.5.4. For this reason, the function

$$z_t'' = \alpha_1 + \alpha_2 t + s_t$$

is plotted in Figure 9.2, for $\alpha_1 = 2.00$, $\alpha_2 = 0.01$, $s_1 = 0.10$, $s_2 = -0.05$, $s_3 = -0.15$, $s_4 = -0.30$, $s_5 = -0.35$, $s_6 = -0.25$, $s_7 = -0.10$, $s_8 = 0.05$, $s_9 = 0.20$, $s_{10} = 0.30$, $s_{11} = 0.35$ and $s_{12} = 0.20$.

9.7.2 The A.R.I.M.A. Model of Order $(1,1,0) \times (0,1,1)_{12}$

The general solution of the A.R.I.M.A. model of order $(1,1,0) \times (0,1,1)_{12}$ is, using equations (9.7.3) and (9.7.8),

$$z_t = \alpha_1 + \alpha_2 t + \alpha_3 \phi_1^t + s_t + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots + \psi_{t-1} a_1 \qquad 9.7.10$$

for $t > 12$. s_t is as defined in Section 9.7.1 and the ψ weights satisfy

$$(1-\phi_1 B)(1-B)(1-B^{12})\, a_t = (1-\theta_1 B^{12})\, a_t$$

The complementary function is composed of the linear trend and seasonal components present in the expression (9.7.9) plus the term $\alpha_3 \phi_1^t$. Unless ϕ_1 is very close to ± 1, the effect of $\alpha_3 \phi_1^t$ diminishes rapidly as t increases and the deterministic component is soon identical to that shown in Figure 9.2 for the model of order $(0,1,1) \times (0,1,1)_{12}$ when the values of α_1, α_2 and s_j $(j = 1,2,3,\ldots,12)$ quoted in Section 9.7.1 are employed.

9.7.3 The A.R.I.M.A. Model of Order $(0,1,1) \times (1,1,0)_{12}$

The A.R.I.M.A. model of order $(0,1,1) \times (1,1,0)_{12}$ has the general solution,

$$z_t = \alpha_1 + \alpha_2 t + \alpha_3 |\Phi_1|^{t/12} + s_t + |\Phi_1|^{t/12} s'_t + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots + \psi_{t-1} a_1 \qquad 9.7.11$$

for $t > 1$. Again, s_t is as defined in Section 9.7.1 while s'_t is given by

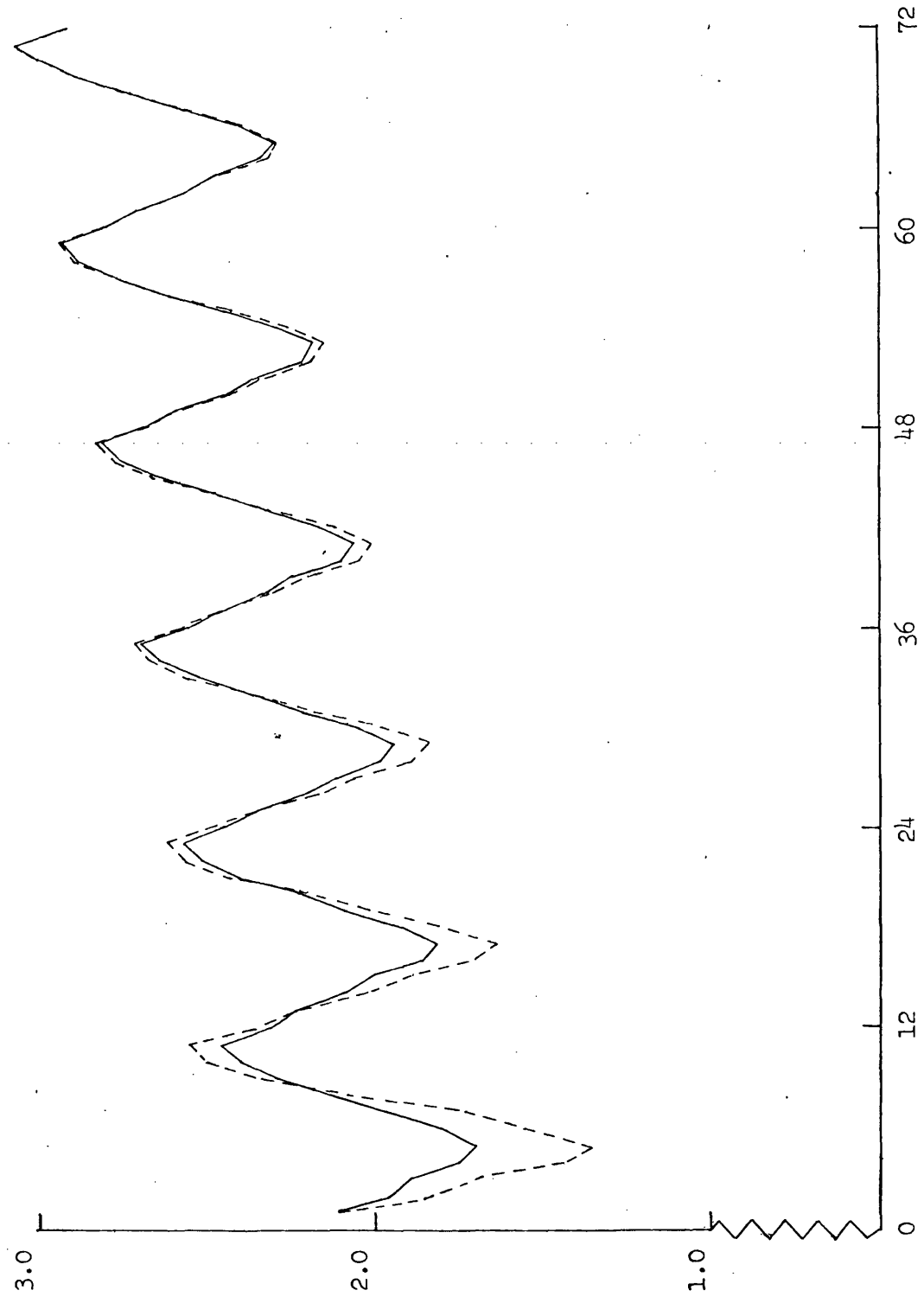


FIGURE 9.2

Examples of complementary functions associated with the models

$(1-B)(1-B^{12})z_t = (1-\theta_1 B)(1-\theta_1 B^{12})a_t$ ——— ; $(1-\phi_1 B^{12})(1-B)(1-B^{12})z_t = (1-\theta_1 B)a_t$ ----.

$$s'_t = (-1)^t \left\{ \frac{c'_o}{2} + \sum_{\nu=1}^{5} \left(c'_\nu \cos \frac{2\pi\nu t}{12} + d'_\nu \sin \frac{2\pi\nu t}{12} \right) \right\}$$

The ψ weights satisfy the equation

$$(1-\Phi_1 B^{12})(1-B)(1-B^{12})\ \psi(B)\ a_t = (1-\theta_1 B)\ a_t$$

The complementary function for this model thus contains the terms $\alpha_3 |\Phi_1|^{t/12}$ and $|\Phi_1|^{t/12}\ s'_t$ over and above the linear trend and additive seasonal components. The term $\alpha_3\ |\Phi_1|^{t/12}$ will generally decay rather slowly, even for values of Φ_1 not particularly close to $\pm$ 1, while $|\Phi_1|^{t/12}\ s'_t$ represents a similarly decaying seasonal variation. The effect of the expressions $\alpha_3\ |\Phi_1|^{t/12}$ and $|\Phi_1|^{t/12}\ s'_t$ obviously depends on the values of A_3 and s'_j ($j = 1,2,3,\ldots,12$) but one possible behaviour is illustrated in Figure 9.2 where the complementary function is plotted for the values of α_1, α_2 and s_j ($j = 1,2,3,\ldots,12$) employed for the model of order $(0,1,1) \times (0,1,1)_{12}$ and $\alpha_3 = 0.1$, $\Phi_1 = \frac{1}{2}$, $s'_j = s_j$ ($j = 1,2,3,\ldots,12$). The complementary function is seen to possess a seasonal component, the amplitude of which decreases as t increases, eventually settling down to the pattern followed by the complementary function of the model of order $(0,1,1) \times (0,1,1)_{12}$.

9.7.4 The A.R.I.M.A. Model of Order $(1,1,0) \times (1,1,0)_{12}$

In the case of the A.R.I.M.A. model of order $(1,1,0) \times (1,1,0)_{12}$, the general solution for $t > 0$ is

$$z_t = \alpha_1 + \alpha_2 t + \alpha_3\ \phi_1^t + \alpha_4\ |\Phi_1|^{t/12} + s_t + |\Phi_1|^{t/12}\ s'_t$$
$$+ a_t + \psi_1\ a_{t-1} + \psi_2\ a_{t-2} + \ldots + \psi_{t-1}\ a_1 \qquad 9.7.12$$

where s_t and s'_t have been defined previously and the ψ weights satisfy

$$(1-\phi_1 B)(1-\Phi_1 B^{12})(1-B)(1-B^{12})\ \psi\ (B) a_t = a_t$$

As in Section 9.7.2, the effect of the term $\alpha_3\ \phi_1^t$ soon dies out

as t increases and apart from the first few values of t, the complementary function is the same as that for the model of order $(0,1,1) \times (1,1,0)_{12}$, an example of which is given in Figure 9.2.

9.7.5 General Comments

In cases where the influence of the random component is small, the complementary function will tend to dominate the general solution of the model in question. When this occurs, models associated with almost the same complementary function would be expected to behave in a similar manner. Thus for example the models discussed in Sections (9.7.1) and (9.7.2) would generate forecasts which agree closely with each other, while the models of order $(0,1,1) \times (0,1,1)_{12}$ $(0,1,1) \times (1,1,0)_{12}$ may well produce quite different forecasts since as we see from Figure 9.2 the complementary functions can differ quite considerably. Reference to Section 3.7 emphasises this point.

When the random component possesses a high variation compared with the other components in the model, the particular integral will dominate the general solution and as t increases the observations will diverge quite quickly from the path predicted by the complementary function evaluated at time t = 0.

9.8 Conclusions

The main object of this chapter has been to gain a better understanding of A.R.I.M.A. models by expressing them in a form which is generally more familiar to most statisticians. We have considered the interpretation of the forecasts generated by A.R.I.M.A. models in terms of E.W.M.A.'s and also the representation of A.R.I.M.A. processes by trend and seasonal models.

In Sections 9.3, 9.4 and 9.5 we found that provided the moving average operator in an A.R.I.M.A. model possessed real roots H_j

lying in the region $0 < H_j < 1$, then it was possible to express the generated forecasts as E.W.M.A.'s. The complexity of these expressions depended on a number of factors. For non-seasonal models the one step ahead forecasts could be described by a quite elegant combination of various orders of E.W.M.A.'s, even in the general case. When the model under consideration included few parameters, the one step ahead forecasts were related to the forecasts produced by techniques based on exponential smoothing. In particular the forecasts resulting from the A.R.I.M.A. models of order (0,1,1), (0,2,2) and (0,3,3) were found to be more general forms of those arrived at by Brown's method, when the local models were respectively constant linear and quadratic. However for A.R.I.M.A. models which included many parameters, the interpretation of higher order E.W.M.A.'s was in itself rather difficult and so the understanding was not improved.

The one step ahead forecast derived from the general multiplicative seasonal model was an unwieldy combination of various orders of E.W.M.A.'s. The forecasts for simpler seasonal models were however arrived at by substituting in the general expression and it was then possible to relate these forecasts to those involving more familiar arguments.

It may be rather misleading to say that the forecasts generated by A.R.I.M.A. models can sometimes be expressed as E.W.M.A.'s. The latter assume an infinite amount of data and although a finite approximation is generally satisfactory for non seasonal models, the same may not be true in the case of seasonal models with moving average seasonal parameters, even for moderate length series.

An alternative interpretation of A.R.I.M.A. models was achieved by solving the difference equations which constitute the models. The general solution consisted of a deterministic component (the complementary function) and an unpredictable component (a particular

integral). For seasonal models based on the differencing operator $\nabla\nabla_{12}$, the complementary function always included a linear trend and an additive seasonal component described by damped sine waves. The other deterministic terms depended on the autoregressive parameters present in the A.R.I.M.A. model, while the effect of these terms was more pronounced for seasonal parameters than non-seasonal parameters. The complementary function tended to dominate the general solution when there was little random variation and the particular integral when the random variation was high.

CHAPTER 10

FURTHER EXPERIENCE WITH THE BOX-JENKINS PROCEDURE

10.1 Introduction

It would be unfair to judge the success or failure of the Box-Jenkins forecasting procedure solely on the basis of the single case study described in Chapters 3 and 4 and without taking into account the performance of other forecasting techniques on the same data. In order to gain more experience with the Box-Jenkins procedure it was applied to a further 4 seasonal time series. A summary of these analyses will be given in Section 10.2 with particular attention being paid to some of the points raised in preceeding chapters.

The Holt-Winters method (see Winters (1960), Coutie et al (1964) and Section 1.2.2) was also applied to the 4 series mentioned above and to the Company X data. The forecasts generated by this method were compared with those resulting from the Box-Jenkins procedure and the relative performance of the two techniques will be reported in Section 10.3. These results will be related to the conclusions arrived at by Reid (1969).

The contents of this chapter will be summarised in Section 10.4.

10.2 A Summary of Some Further Box-Jenkins Analyses

We begin this section by briefly describing the 4 seasonal time series on which the Box-Jenkins procedure was employed. These series will be referred to as series I, II, III and IV.

SERIES I

This series is composed of the 4-weekly sales figures of a food product. The 57 observations available are given in Appendix I and plotted in Figure 10.1. A high seasonal variation (period 13) is

apparent while a slight downward trend can be detected.

SERIES II

This series consists of the monthly retail passenger car sales in the U.S.A. The data have been tabulated by Makridakis and Wheelwright (1972) who also give the results (without details) of a Box-Jenkins analysis. The series, which is shown in Figure 10.2, is influenced by strong seasonal and cyclical factors and is composed of 84 observations.

SERIES III

This series is a scaled* approximation to the telephone time series analysed by Tomasek (1972). The data, composed of 108 monthly observations (tabulated in Appendix I and graphed in Figure 10.3), possess an extremely high seasonal variation and a significant upward trend.

SERIES IV

Series IV consists of quarterly despatch figures in the footwear industry. The series shows a fairly high seasonal variation and a cyclical factor. The 48 observations which make up the series are tabulated in Appendix I and plotted in Figure 10.4.

In applying the Box-Jenkins procedure, each series was divided into two parts. The first part was used for fitting the appropriate A.R.I.M.A. model and the second part for examining the forecasting performance of the fitted model. Obviously it was necessary for both the fitting and forecasting periods to be of a reasonable length. Table 10.1 gives the length of these periods chosen for the series I to IV.

---

* Unfortunately the author does not tabulate the data and a graph of the series does not include any scale. Hence series III was obtained by a suitable choice of scale.

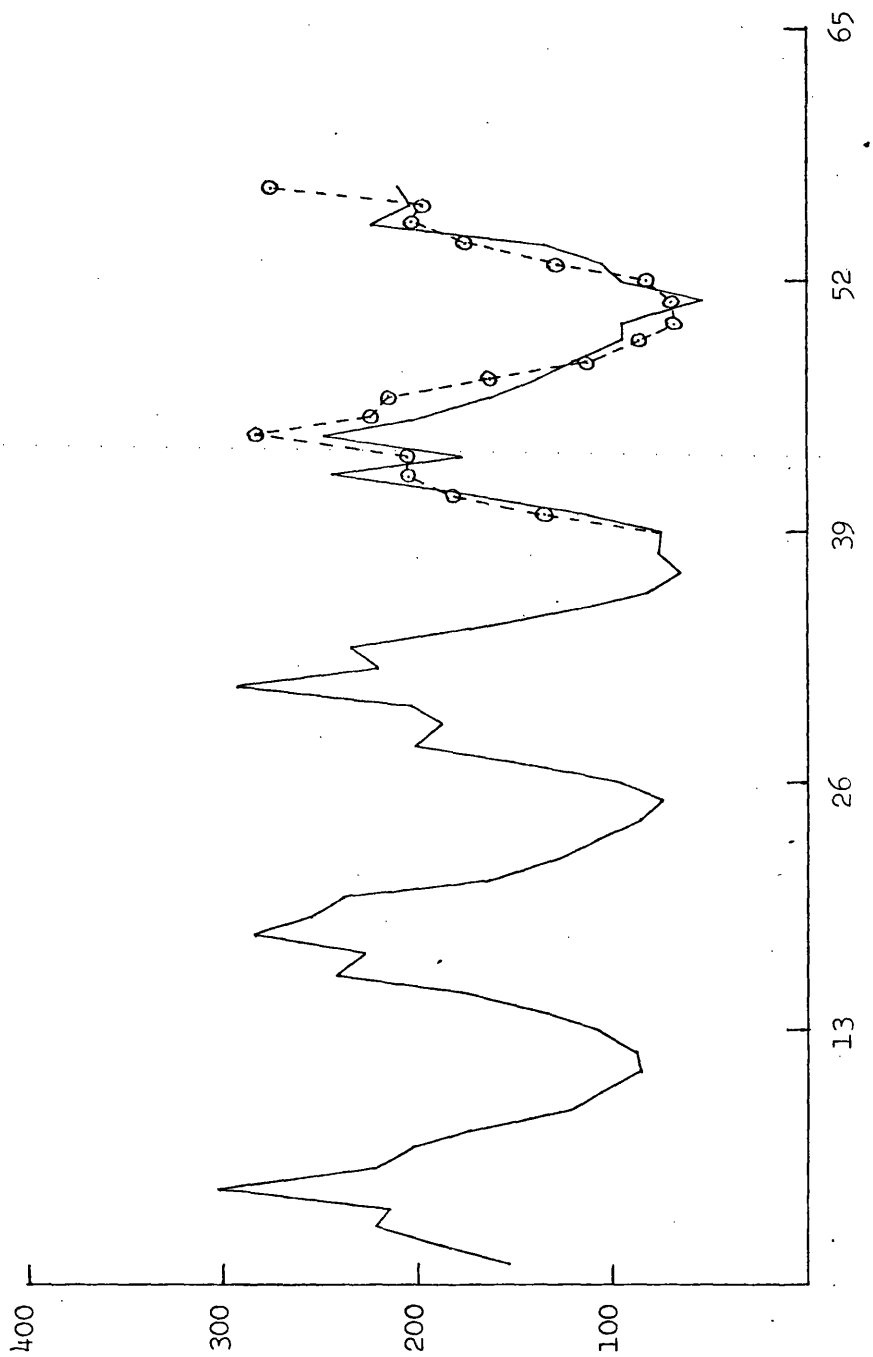


FIGURE 10.1

Series I and 1 step ahead forecasts generated after observation 39.

Actual ———; forecast 0--0--0.

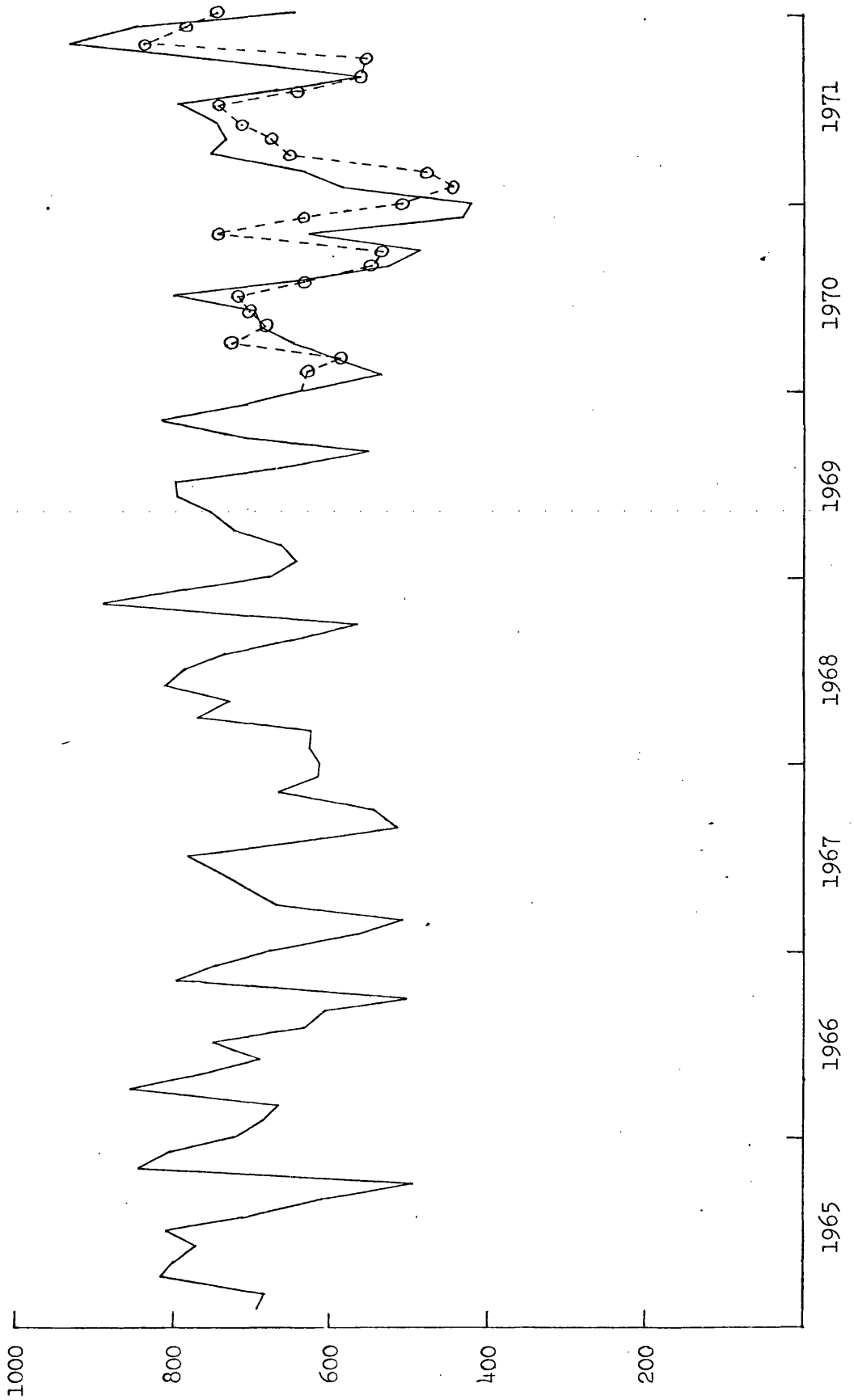


FIGURE 10.2

Series II and 1 step ahead forecasts generated after observation 60

Actual ———; forecast ⊙--⊙--⊙.

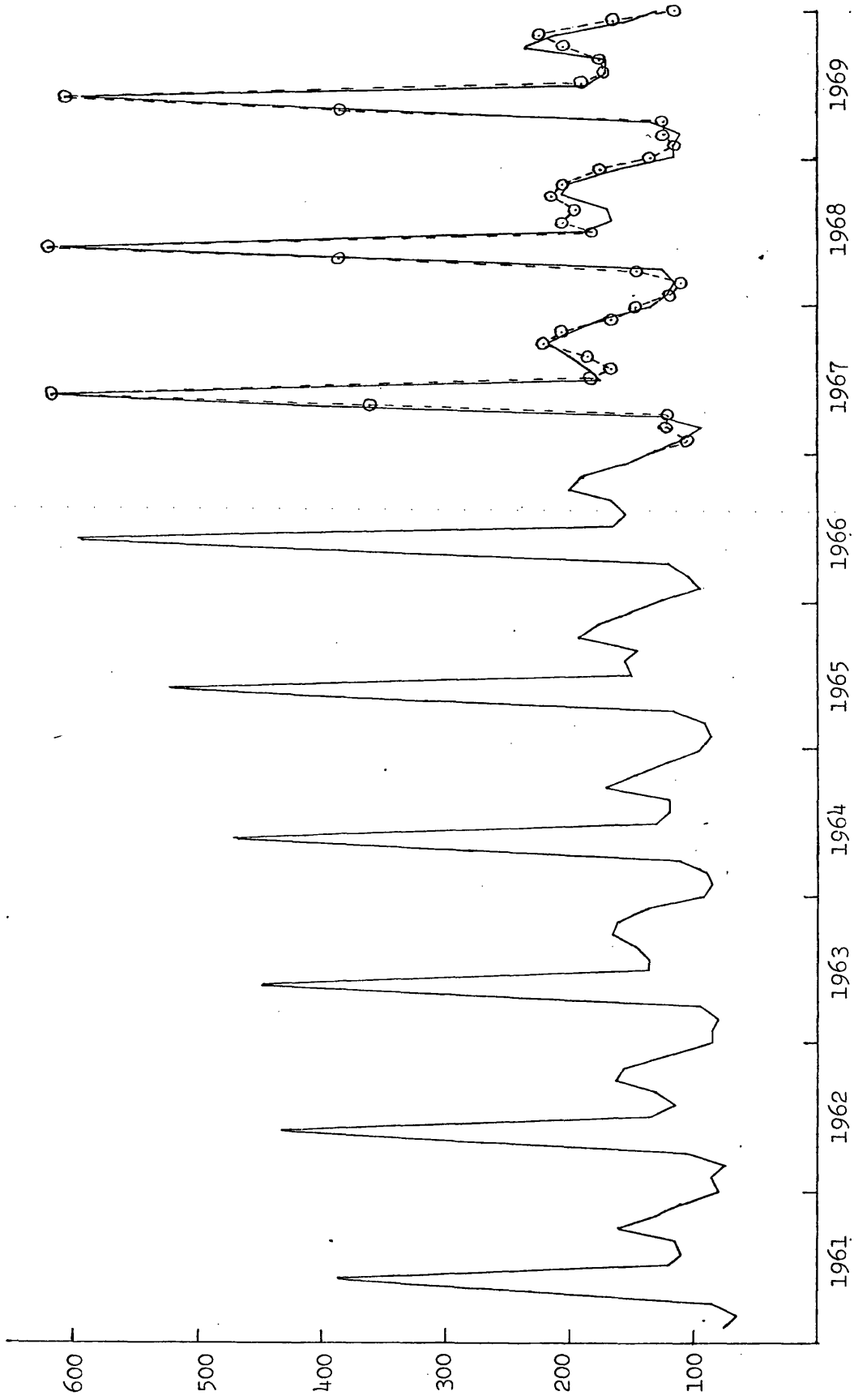


FIGURE 10.3

Series III and 1 step ahead forecasts generated after observation 72

Actual ———; forecast 0--0--0.

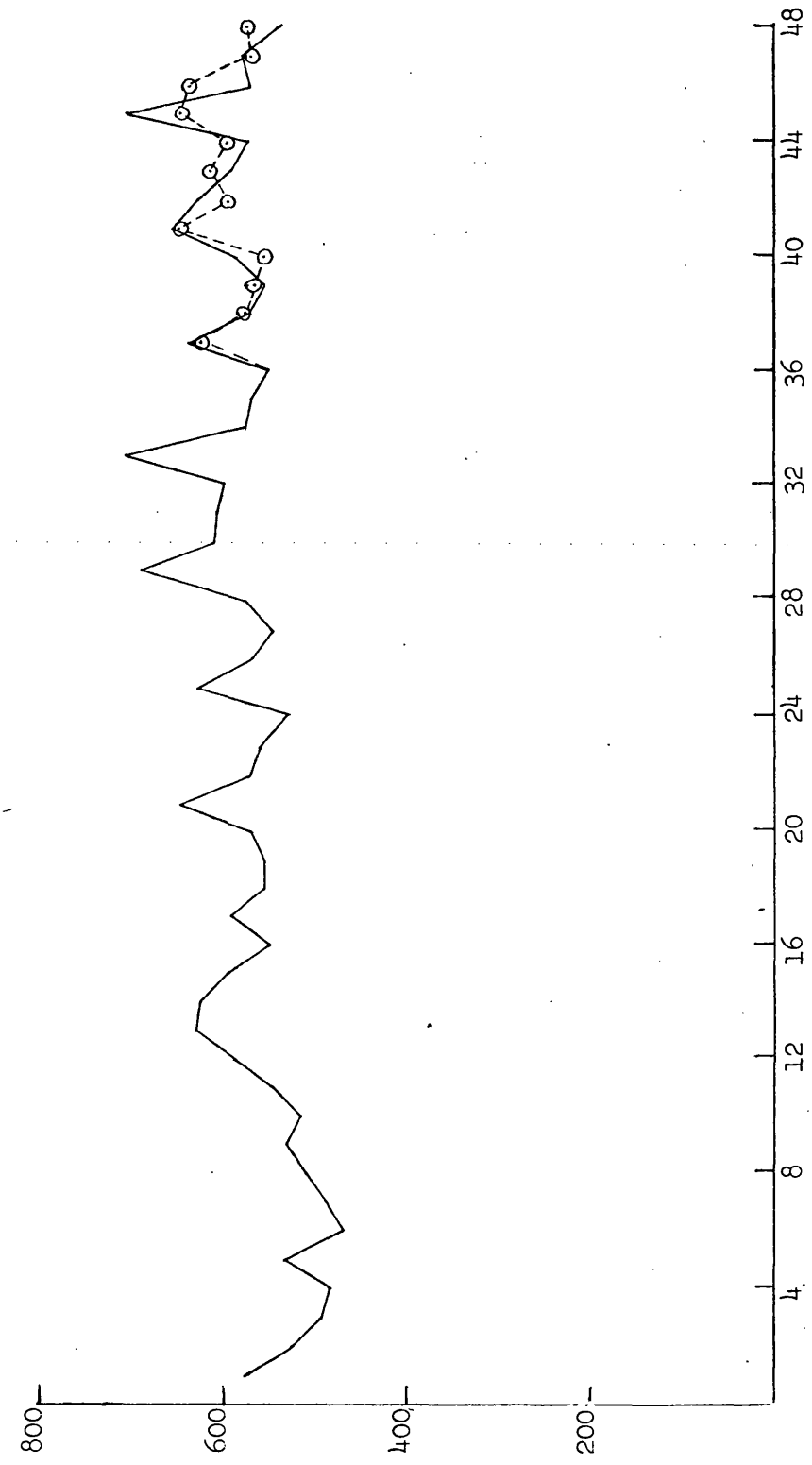


FIGURE 10.4

Series IV and 1 step ahead forecasts generated after observation 36

Actual ———; forecast 0--0--0

Table 10.1 Fitting periods and forecasting periods for series I to IV

Series	Total Length	Fitting Period	Forecasting Period
I	57	39	18
II	84	60	24
III	108	72	36
IV	48	36	12

Using the iterative process of identification, estimation and diagnostic checking described in Chapter 2, the following fitted models were selected to describe the series I to IV.

SERIES I

$$w_t = \nabla_{13} X_t = (1 - 0.95B^{13})a_t - 4.62 \qquad 10.2.1$$

and $\hat{\sigma}_a^2 = 205.9$

SERIES II

$$w_t = \nabla\nabla_{12} X_t = (1 - 0.69B)\ (1 - 0.94B^{12})a_t \qquad 10.2.2$$

and $\hat{\sigma}_a^2 = 3179$

SERIES III

$$w_t = \nabla\nabla_{12} X_t = (1 - 0.47B - 0.49B^2)a_t \qquad 10.2.3$$

and $\hat{\sigma}_a^2 = 217.6$

SERIES IV

$$w_t = \nabla\nabla_4 X_t = (1 - 0.14B)\ (1 - 0.96B^4)a_t \qquad 10.2.4$$

and $\hat{\sigma}_a^2 = 8863719$

In each case X_t denotes the untransformed observation at time t.

Throughout the model fitting procedures a number of the points raised by the analysis of the Company X data were again encountered. These will now be discussed.

It is recalled from Chapters 3 and 4 that a non-linear transformation was applied to the Company X data prior to the Box-Jenkins analysis. The monthly growth rates for the series I to IV are much smaller than that associated with the Company X data making it less important whether a model with additive or multiplicative seasonal effects is assumed. After inspecting the Figures 10.1 to 10.4 we decided to analyse the untransformed observations in each case and thus the extra problems involved in estimating the transformation parameter (see Section 4.4) were avoided.

The importance of the role played by the differencing operator in the A.R.I.M.A. model (discussed in Chapter 5) was emphasised during the fitting of the four models (10.2.1) to (10.2.4). Table 10.2 quotes the estimated variances of the original observations X_t, the differenced series w_t and the residuals a_t resulting from the fitted models (10.2.1) to (10.2.4). All estimates were computed over the fitting period.

Table 10.2 Estimated variances of X_t, w_t and a_t

Series	$\hat{\sigma}^2_x$	$\hat{\sigma}^2_w$	$\hat{\sigma}^2_a$
I	4,620	432.3	205.9
II	8,885	7488	3179
III	11,949	278.4	217.6
IV	27,600,000	12,167,948	8,863,719

If the percentage of the total variation unexplained by each fitted model is taken to be $\hat{\sigma}^2_a/\hat{\sigma}^2_x$ x 100, the percentage variation accounted for by the differencing operation to be $(1 - \hat{\sigma}^2_w/\hat{\sigma}^2_x)$ x 100 and the percentage variation explained by fitting the A.R.M.A. model to w_t to be $(\hat{\sigma}^2_w/\hat{\sigma}^2_x - \hat{\sigma}^2_a/\hat{\sigma}^2_x)$ x 100, then the results given in Table 10.2 are best interpreted in the manner shown below.

Table 10.3 Variation explained by differencing and fitting the A.R.M.A. Model

Series	Percentage of Total Variation Explained		
	By Differencing	By fitting the A.R.M.A. Model	Unexplained
I	90.6	4.9	4.5
II	15.7	48.5	35.8
III	97.7	0.5	1.8
IV	55.9	12.0	32.1

For the two series I and III,the percentage of the total variation unexplained by the fitted models (10.2.1) and (10.2.3) is very small compared with other variations in the series (i.e. due to trend and seasonality) and in both instances most of the total variation is explained by the differencing operation. In contrast, the fitting of the A.R.M.A. model to the stationary series accounts for only a small percentage of the total variation, thus supporting the view of Akaike (1973). The variation unexplained by the models (10.2.2) and (10.2.4) when fitted to the series II and IV represents a much higher percentage of the total variation in these series. In the case of series IV, the effectiveness of the A.R.I.M.A. model is again determined mainly by the differencing operation but for series II the fitting of the A.R.M.A. model explains much more of the total variation than the differencing operation,

As we can see from equations (10.2.1), (10.2.2) and (10.2.4),

the A.R.I.M.A. models fitted to the series I, II and IV all included a moving average seasonal parameter close to unity. Hence, for the reasons given in Chapter 6, one iterative cycle was not sufficient when computing the unconditional sum of squares corresponding to the estimates of these parameters. In fact, we found that at least 4 iterations were necessary in all three cases. Although not immediately obvious, the fitted model (10.2.3) is also close to the non-invertibility boundary since the moving average operator can be expressed as

$$(1 - 0.47B - 0.49B^2)a_t = (1 - 0.97B)\ (1 + 0.50B)a_t \qquad 10.2.5$$

However in this case one iterative cycle was sufficient because the parameter in question is non-seasonal and also the fitting period for the series III is of moderate length.

As in the analysis of the Company X data, we discovered that it was possible to fit several models to each set of data without any inadequacies being detected by the diagnostic checks. For example, no discrepancies were revealed when the model

$$w_t = \nabla_{13}X_t = a_t - 4.62 \qquad 10.2.6$$

was fitted to the series I, although the model (10.2.1) was chosen because it gave rise to a much smaller estimate for σ_a^2. It was however often difficult to decide whether one model fitted the data significantly better than some alternative model. The fact that our series were rather on the short side undoubtedly contributed towards this problem but nevertheless it serves to emphasise that one can never be absolutely sure that any one particular model is appropriate.

Forecasts were generated by the model (10.2.1) (up to lead time 13), models (10.22) and(10.2.3) (up to lead time 12) and model (10.2.4) (up to lead time 4), over the second part of the series in question. The one step ahead forecasts in each case are plotted in Figures 10.1 to 10.4. Although some bias is present in the forecast errors it was

not found to be statistically significant for any of the four series.

The mean squared forecast errors (S.M.S.E (ℓ)) for the lead times mentioned above were computed and are given in Table 10.4. Also tabulated are the corresponding theoretical mean squared errors (T.M.S.E (ℓ)) obtained using the formula (5.5.6).

It can be seen that the agreement between the two quantities is quite reasonable for the mdoels (10.2.1) and (10.2.3), fitted to the series I and III which exhibited small random variations, but not so good for the other two fitted models. However, in the case of the series IV, the forecasting period is so short (only 12 observations) that the results are bound to be rather unreliable.

We concluded in Chapter 9 that the one step ahead forecasts generated by seasonal A.R.I.M.A. models can be expressed in terms of E.W.M.A.'s, provided the roots H_j and J_i of the moving average operator are real and satisfy $0 < H_j < 1$, $0 < J_i < 1$. The expressions for the fitted models (10.2.2) and (10.2.4) are arrived at by substituting $A_1 = 1-\theta_1 = 0.31$, $C_1 = 1-\theta_1 = 0.06$ and $A_1 = 0.86$, $C_1 = 0.04$ respectively in equation (9.5.11), remembering that in the latter case the seasonal period is 4 and not 12. The one step ahead forecast derived from the fitted model (10.2.1) is apart from a constant, simply the yearly E.W.M.A. $\bar{z}^{(0.05)}_{t-11}$. However the interpretation of each of the three models mentioned above is affected by the practical consideration discussed in Section 9.6. It is not possible to represent the forecast generated by the fitted model (10.2.3) entirely in terms of E.W.M.A.'s since one of the roots H_j is negative (see equation (10.2.5)).

A further means of interpreting A.R.I.M.A. models, dealt with in Chapter 9, involves the solution of the appropriate difference equation. On applying this approach to the models (10.2.1) to (10.2.4) we find that the complementary functions consist of a linear trend and an additive seasonal component, except for the model (10.2.1) when no linear term is present. Thus as we saw in Section 7.2, the models could have

Table 10.4 Actual and theoretical M.S.E.'s. for the 4 fitted models

Lead Time	Series I		Series II		Series III		Series IV	
	S.M.S.E.(ℓ)	T.M.S.E.(ℓ)	S.M.S.E.(ℓ)	T.M.S.E.(ℓ)	S.M.S.E.(ℓ)	T.M.S.E.(ℓ)	S.M.S.E.(ℓ)	T.M.S.E.(ℓ)
1	905.7	905.7	8732	8732	195.8	195.8	9480000	9480000
2	925.2	905.7	12355	9570	289.1	250.9	8193373	16495200
3	975.7	905.7	15385	10409	284.6	251.1	13387181	23510200
4	944.2	905.7	16391	11247	298.6	251.4	12691567	30525200
5	960.9	905.7	17644	12085	302.8	251.7	-	-
6	938.2	905.7	18683	12923	314.8	252.0	-	-
7	980.5	905.7	20549	13762	330.4	252.3	-	-
8	814.9	905.7	24099	14600	338.0	252.6	-	-
9	847.7	905.7	29080	15438	336.8	252.9	-	-
10	937.2	905.7	32980	16276	353.4	253.2	-	-
11	1046.4	905.7	32898	17114	371.7	253.5	-	-
12	1101.1	905.7	25832	17952	389.9	253.8	-	-
13	1237.0	905.7	-	-	-	-	-	-

arisen from the sum of a deterministic component and a non-stationary error term. For example, since the moving average seasonal parameter in the model (10.2.2) is close to unity the series X_t could have been generated by

$$X_t = \alpha + \beta t + s_t + n_t$$

where $\nabla n_t = a_t - \theta a_{t-1}$

and α, β and s_j $(j = 1,2,3,\ldots,12)$ are constants with $s_t = s_{t-12}$.

However, since there is no physical reason for assuming that a totally deterministic component is present in any of the series I to IV it is perhaps more likely that the models (10.2.1) to (10.2.4) result from processes of the type examined in Chapter 8.

10.3 Comparison with the Holt-Winters Method

The Holt-Winters method described by Winters (1960) was applied to the series I to IV and also to the Company X data. For the series I to IV the fitting and forecasting periods employed were the same as those for the Box-Jenkins analyses (see Table 10.1) while for the Company X data the fitting and forecasting periods were composed of 60 and 17 observations respectively. The fitting periods were used for estimating smoothing constants and starting values for the trend and seasonal factors. Forecasts were then generated over the second part of each series.

The forecasting performances of the Box-Jenkins and Holt-Winters procedures were compared on the basis of the mean absolute forecast errors over the second part of the series under consideration. The use of this statistic assumes a cost of error function of the form

$$C(e) = a|e| \qquad \text{for } a > 0$$

which differs from the quadratic cost of error function employed by

Reid (1969, 1971) and Newbold and Granger (1974). However, since our interest was in the comparative performance of two forecasting techniques the choice of cost function is not a crucial one (see Granger and Newbold (1973)). In fact our conclusions would have been exactly the same had a quadratic cost of error function been assumed, as remarked by Reid (1971).

Table 10.5 gives the mean absolute one step ahead forecast errors arising from the use of the Box-Jenkins and Holt-Winters techniques.

Table 10.5 Mean absolute one step ahead forecast errors

Series	Mean Absolute Forecast Errors	
	Box-Jenkins	Holt-Winters
I	25.5	20.9
II	73.3	73.5
III	11.3	10.3
IV	2481	2522
Company X	50.9	44.9

It should be pointed out that the A.R.I.M.A. model fitted to the Company X data was the model (4.3.1) and the mean absolute forecast error was computed after re-transforming the forecasts.

From Table 10.5 we see that Box-Jenkins produced the smaller mean absolute forecast errors for series II and IV and Holt-Winters for series I, III and the Company X series. Using Reid's decision tree (see Kendall (1973, page 127)) one would have expected Box-Jenkins to have performed better than Holt-Winters on the series II, III and Company X (all with fitting period > 50 observations) with Holt-Winters doing better on the series I and IV ($\leq$ 50 observations). Of course Reid's diagram is based on generalisations from a finite sample and as he admits "a good deal of "hunch"" is embodied in the diagram.

It is of interest to note that for the two series II and IV in which Box-Jenkins does better than Holt-Winters, we see from Table 10.3 that the random (unpredictable) component possesses a high variance compared with other variability in these series.

We now look briefly at the mean absolute errors resulting from forecasts for higher lead times. Table 10.6 quotes the mean absolute lead time 6 forecast errors for the series I to III and Company X and the mean absolute lead time 4 forecast errors for series IV.

Table 10.6 Mean absolute lead time 6 forecast errors, series I, II II and Company X, mean absolute lead time 4 forecast errors, series IV

Series	Mean Absolute Forecast Errors	
	Box-Jenkins	Holt-Winters
I	24.9	17.9
II	110.3	111.4
III	14.0	11.8
IV	3227	3117
Company X	77.3	52.3

The only series for which Box-Jenkins performs better than Holt-Winters is series II. In fact it would appear that Holt-Winters compares more favourably with Box-Jenkins for higher lead times than for the lead time 1 forecasts.

10.4 Summary

In this chapter, the Box-Jenkins forecasting procedure has been applied to a further 4 seasonal time series. A number of the points raised by the analysis of the Company X data and dealt with in Chapters 5 to 9

were again apparent. These included the importance of the choice of differencing operator, the possibility of several A.R.I.M.A. models fitting the data almost equally well, the estimation of moving average seasonal parameters and the interpretation of A.R.I.M.A. models and the forecasts generated by such models. However, since non-linear transformations of the data were not deemed to be necessary, the problems of estimating a transformation parameter by the methods described in Chapter 4 were not encountered.

The performance of the Box-Jenkins procedure was compared with that of the Holt-Winters method on the 4 series introduced in Section 10.2 and also on the Company X data. Overall, the Holt-Winters method compared most favourably with the Box-Jenkins procedure for both lead time 1 forecasts and forecasts for higher lead times. However, it would be unwise to attempt to draw any general conclusions on the basis of these results, for the following reasons:

(1) The comparison was performed on a sample of only 5 series. The conclusions arrived at from the comprehensive studies conducted by Reid (1969) and Newbold and Granger (1974) (see Section 1.4) suggest that the Box-Jenkins procedure generally performs better than the Holt-Winters method, even on seasonal data.

(2) The fitting and forecasting periods (series III apart) were rather short.

(3) The number of subjective decisions involved in the Box-Jenkins procedure makes it extremely likely that different practitioners would fit different A.R.I.M.A. models. It is therefore possible that someone more experienced with the Box-Jenkins procedure could have improved upon the Box-Jenkins results given in Tables 10.5 and 10.6.

CHAPTER 11

SUMMARY, CONCLUSIONS AND FURTHER RESEARCH

11.1 Summary and Conclusions

This thesis has been concerned with an examination of various practical problems related to the use of the forecasting procedure proposed by Box and Jenkins (1970). The emphasis has been placed on the application of this procedure in the particular area of seasonal forecasting.

We began our examination of the Box-Jenkins procedure by applying it to a seasonal series consisting of the monthly sales figures of an engineered product. The forecasts generated during this analysis were intuitively very poor while in addition other more general features of the procedure were considered to be worthy of more detailed attention. These points provided the stimulus for the topics examined in Chapters 4 to 9.

The employment of non-linear transformations in time series analyses was considered in Chapter 4. It was shown that for a series which possesses an approximate linear trend and multiplicative seasonal variation (e.g. the Company X data), a logarithmic trnasformation can be justified as long as the monthly growth rate is small. When this is not the case one approach is to assume a general non-linear transformation of the kind discussed by Box and Cox (1964). However, the estimation of the transformation parameter is a long, complicated exercise which requires the prior identification of an A.R.I.M.A. model. Further it may be necessary to perform this estimation procedure at frequent intervals since we saw that the value of the transformation parameter can change quite significantly over short periods of time. In view of this possibility it would seem advisable to avoid a non-linear

transformation wherever possible by applying a suitable differencing operator to the untransformed observations.

Chapter 5 dealt with the important problem of selecting the degree of differencing which reduces a non-stationary series to some stationary process. It was proposed that the decomposition of a series into trend, seasonal and error components and the theoretical behaviour of the variance of a stationary series could both be used, in addition to the sample autocorrelation function, when identifying the differencing operator. We illustrated just how crucial this identification can be by quoting an example in which two A.R.I.M.A. models based on different degrees of differencing fitted the transformed Company X data almost equally well, yet the two sets of tolerance limits associated with the generated point forecasts differed considerably, particularly for higher lead times. However, in general, for short or even moderate length series it is often impossible to decide which of two models, fitting the data equally well yet forecasting quite differently, is appropriate.

The estimation stage of the model fitting process in the Box-Jenkins procedure was investigated in Chapter 6. When analysing seasonal (especially monthly) data we discovered that an already complicated estimation procedure can become even more involved when the identified model includes a moving average seasonal parameter close to the non-invertibility boundary. In such a case several iterative cycles are usually necessary in computing the unconditional sum of squares for parameter values in the neighbourhood of the estimated value, even for moderate length series. A similar problem may also occur when estimating moving average parameters for short non-seasonal series.

In Chapter 7 we showed that the estimation situation discussed in Chapter 6 can occur when fitting A.R.I.M.A. models to series which include deterministic components. This theory was confirmed when the

Box-Jenkins procedure was applied to a short generated series possessing a deterministic linear trend. During this analysis we found that when fitting an A.R.M.A. model to the stationary series which had a non-zero mean, it was important to estimate this parameter by least squares. A simple linear regression on time was also performed on the generated series and on the basis of the two analyses we concluded that although the Box-Jenkins procedure produced the slightly better forecasts, the use of the latter could not be justified in preference to the less expensive linear regression approach.

The theoretical A.R.I.M.A. models resulting from the aggregate of stochastic processes representing trend, seasonal and extraneous error components were derived in Chapter 8. We found that even when the individual components were described by quite simple A.R.I.M.A. models the aggregate process generally involved many parameters although it was shown that in practice it may be possible to approximate by a less complicated model, depending on the relative sizes of the variances of the white noise processes associated with each component.

In Chapter 9 we discussed the interpretation of the forecasts generated by A.R.I.M.A. models, in terms of E.W.M.A.'s. Such an interpretation is generally possible provided the roots of the moving average operators (both seasonal and non-seasonal) are real and positive. The expressions for the one step ahead forecasts are quite simple for models with few parameters but for multi-parameter models (particularly seasonal) we reach the stage where the E.W.M.A.'s are themselves difficult to interpret. One point of practical importance is that for short or moderate length series, the expressions for the forecasts generated by seasonal models will bear little resemblance to E.W.M.A.'s when the model in question includes a moving average seasonal parameter close to the non-invertibility boundary.

An alternative means of interpreting A.R.I.M.A. models was also investigated in Chapter 9. The difference equation form of the

A.R.I.M.A. model was solved in terms of a deterministic component, representing trend and seasonality, and an unpredictable component. The deterministic component consists of polynomials and cyclical terms (arising from the differencing operator in the A.R.I.M.A. model) and damped exponentials and damped sine waves (arising from the stationary autoregressive operators).

The application of the Box-Jenkins procedure to a further 4 seasonal time series was described in Chapter 10. Many of the points raised by the analysis of Chapter 3 and dealt with in Chapters 4 to 9 were again encountered. In addition, the performance of the Box-Jenkins procedure was compared with that of the Holt-Winters method on the 4 series mentioned above and also on the Company X data. The latter method compared most favourably with the Box-Jenkins procedure although it would be unwise to draw any general conclusions on the basis of 5 rather short series. However this comparison did support the view of several authors (e.g. Reid (1969, 1971), Chatfield and Prothero (1973 a), Newbold and Granger (1974)) that the main disadvantage of the Box-Jenkins procedure compared with most other univariate techniques are

i) the expense involved in fitting an initial model

ii) it can only be used effectively by an experienced skillful statistician

and iii) it requires rather more data than are often available.

The Box-Jenkins procedure certainly proved more expensive in terms of computing time than the Holt-Winters method and this was especially true in the case of the Company X data when a non-linear transformation was performed for the Box-Jenkins analysis whereas no such transformation was necessary when applying the Holt-Winters technique. The need for experience and skill in the Box-Jenkins procedure is brought about mainly by the number of subjective decisions involved in the identification of an A.R.I.M.A. model although as we saw in Chapter 4 the aforementioned

qualities are equally essential in deciding whether any transformation should be applied to the data. Again, until a more general understanding of A.R.I.M.A. models is achieved it may be necessary to relate these models to the more familiar concepts of trend and seasonality, an operation which, as demonstrated in Chapter 9, can itself be quite complicated. The point that the Box-Jenkins procedure requires rather more data than are often available was emphasised by the problems arising when short series were analysed. In such cases the tools employed in the identification process are unreliable, the estimation situation discussed in Chapter 6 is more likely to occur while the diagnostic checks become rather insensitive, making it often impossible to decide between two or more A.R.I.M.A. models which apparently fit the data equally well yet forecast quite differently.

Against these practical difficulties, the results of Reid (1969, 1971) and Newbold and Granger (1974) suggest that when used by an experienced practitioner, the Box-Jenkins procedure generally outperforms the other invariate forecasting techniques reviewed in Chapter 1. This is perhaps not surprising in view of the statement of Box and Jenkins (1973) that"the alternative and traditional commonsense forecasting methods ..... are for most part special cases of the A.R.I.M.A. model". However this statement should not be taken to mean that other forecasting methods should never be used. The important point is whether the extra complication and expense involved in the Box-Jenkins procedure can be justified. Reid (1971) suggests that the additional expense may be justified for macro-economic series but not for sales forecasting when a large number of individual items are involved. In the latter case the fully automatic method of Holt-Winters would probably be preferred (see Chatfield and Prothero (1973 a)).

Only experience in the form of more comparative studies, similar to those performed by Reid (1969) and Newbold and Granger (1974), will reveal the conditions under which the models assumed by automatic

forecasting methods are a good enough approximation to the "true" underlying process. For the 5 seasonal series considered in Chapter 10, the extra cost and complexity of the Box-Jenkins procedure could not be justified in preference to the method of Holt-Winters.

11.2 Further Research

Throughout this thesis we have concentrated on the performance of the Box-Jenkins procedure when generating forecasts based solely on past values of the variable to be forecasted (i.e. used as a univariate procedure). This procedure can however be extended to produce forecasts based not only on past values of the series being forecasted but also on past values of other related series (i.e. a multivariate procedure). Box and Jenkins (1970, Part III) have introduced a class of models, called transfer function models, which relate on "output" variable to one or more "input" variables. These models can be expressed in a similar notation to the A.R.I.M.A. models encountered throughout this thesis but the model fitting process is naturally much more involved than in the univariate case. The generation of forecasts employing such models is termed "forecasting using leading indicators". Apart from the additional complications associated with fitting a transfer function model, there is also the problem of selecting suitable related variables. Occasionally the choice may be obvious (e.g. in the case of series IV introduced in Chapter 10, a "stocks" series was supplied together with the "despatches" series, so that the former could be used in forecasting the latter) but generally the input variables need to be carefully selected.

It would be an interesting exercise to seek variables which are related to the series analysed in Chapters 3 and 10 and then to use the Box-Jenkins multivariate procedure for generating forecasts. These forecasts could then be compared with the univariate forecasts already computed. For the series in which only a small percentage

of the total variation remains unexplained by an A.R.I.M.A. model the employment of other variables could not be expected to greatly improve the accuracy of the forecasts, but when a high percentage of the total variation is unexplained by an A.R.I.M.A. model there is certainly room for improvement. However, the Box-Jenkins multivariate procedure is relatively untried and practical improvements are still being investigated (see e.g. Newbold (1973 b)). As yet, apart from the examples given in Box and Jenkins (1970), few case studies appear in the literature and it would seem that much experience is necessary with the univariate procedure before embarking on the multivariate procedure.

ACKNOWLEDGEMENTS

I wish to thank my supervisor Dr. C. Chatfield for his considerable assistance and encouragement during the preparation of this thesis.

I am most grateful to Company X for supplying the data analysed in Chapter 3 and to H.J. Heinz Co. Ltd., Hayes, Middlesex, and Company Y for supplying series I and IV respectively.

Financial support was provided by the Science Research Council.

REFERENCES

Akaike, H. (1973). Contribution to discussion of "Box-Jenkins seasonal forecasting: problems in a case study", by C. Chatfield and D.L. Prothero. J.R. Statist. Soc. A, 136, 330.

Anderson, T.W. (1971). The Statistical Analysis of Time-series. New York: Wiley.

Bartlett, M.S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. J.R. Statist. Soc. Suppl., 8, 27-41.

Bates, J.M. and Granger, C.W.J. (1969). The combination of forecasts. Oper. Res. Q., 20, 451-468.

Beyer, W.H. (1968). Handbook of Tables for Probability and Statistics. Ohio: The Chemical Rubber Co.

Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time-series. Biometrika, 60, 217-226.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. J.R. Statist. Soc. B, 26, 211-252.

Box, G.E.P. and Jenkins, G.M. (1968). Some recent advances in forecasting and control. Part I. Appl. Statist., 17, 91-109.

Box, G.E.P. and Jenkins, G.M. (1970). Time-Series Analysis, forecasting and control. San francisco: Holden-Day.

Box, G.E.P. and Jenkins, G.M. (1973). Some comments on a paper by Chatfield and Prothero and on a review by Kendall. J.R. Statist. Soc. A, 136, 337-345.

Box, G.E.P. and Newbold, P. (1971). Some comments on a paper of Coen, Gomme and Kendall. J.R. Statist. Soc. A, 134, 229-240.

Box, G.E.P. and Pierce, D.A. (1970). Distribution of residual auto-correlation in autoregressive-integrated moving average time-series models. J. Amer. Statist. Assoc., 65, 1509-1526.

Bridge, J.L. (1971). Applied Econometrics. Amsterdam: North Holland Publishing Company.

Brown, R.G. (1963). Smoothing, Forecasting and Prediction of Discrete Time-series. New Jersey: Prentice Hall.

Chambers, J.C., Mullick, S.K. and Smith, D.D. (1971). How to choose the right forecasting technique. Harvard Business Review, 49, No. 4, 45-74.

Chatfield, C. (1974). A review of forecasting techniques. Proceedings of the 1974 E.S.O.M.A.R. conference on "forecasting in marketing".

Chatfield, C. and Prothero, D.L. (1973 a). Box-Jenkins seasonal forecasting: problems in a case study. J.R. Statist. Soc. A, 136, 295-315.

Chatfield, C. and Prothero, D.L. (1973 b). Reply to "Some comments on a paper by Chatfield and Prothero and on a review by Kendall". J.R. Statist. Soc. A, 136, 345-352.

Christ, C.F. (1966). Econometric Models and Methods. New York: Wiley.

Cleveland, W.S. (1972). The inverse autocorrelations of a time-series and their applications. Technometrics, 14, 277-293.

Coen, P.J., Gomme, E.D., and Kendall, M.G. (1969). Lagged relationships in economic forecasting. J.R. Statist. Soc. A, 132, 133-163.

Coutie, G.A., Davies, O.L., Hossell, C.M., Millar, D.W.G.P., and Morrell, A.J.M. (1964). Short-term Forecasting. I.C.I. Monograph No. 2. Edinburgh: Oliver and Boyd.

Dickinson, J.P. (1973). Some statistical results in the combination of forecasts. Oper. Res. Q., 24, 253-260.

Granger, C.W.J. (1972). Time-series modelling and interpretation. Paper presented to the Eur. Econom. Congress in Budapest, Sept. 1972.

Granger, C.W.J. (1973). Multi-step forecasting errors and model mis-specification. Paper presented to the Econom. Soc., European Meeting, August 1973.

Granger, C.W.J. and Newbold, P. (1970). Forecasting transformed variables. Nottingham University Forecasting Project, Note 6.

Granger, C.W.J. and Newbold, P. (1972). Economic forecasting - the atheist's viewpoint. Paper presented to conference on "the Modelling of the U.K. Economy", London Graduate School of Business Studies.

Granger, C.W.J. and Newbold, P. (1973). Some comments on the evaluation of economic forecasts. Appl. Econ., 5, 35-47.

Green, M. and Harrison, P.J. (1973). Fashion forecasting for a mail order company using a Bayesian approach. Oper. Res. Q., 24, 193-205.

Gregg, J.V., Hossell, C.M. and Richardson, J.T. (1964). Mathematical Trend Curves: An Aid to Forecasting. I.C.I. Monograph No. 1. Edinburgh: Oliver and Boyd.

Harrison, P.J. (1965). Short-term sales forecasting. Appl. Statist., 14, 102-139.

Harrison, P.J. and Pearce, S.P. (1972). The use of trend curves as an aid to market forecasting. Industrial Marketing Management, 2, 149-170.

Harrison, P.J. and Stevens, C.F. (1971). A Bayesian approach to short-term forecasting. Oper. Res. Q., 22, 341-362.

Jenkins, G.M. and Watts, D.G. (1968). Spectral Analysis and its Applications. San Francisco: Holden-Day.

Kendall, M.G. (1971). Review of "Time-Series Analysis, forecasting and control" by G.E.P. Box and G.M. Jenkins. J.R. Statist. Soc. A, 134, 450-453.

Kendall, M.G. (1973). Time-series. London: Griffin.

Kendall, M.G. and Stuart, A. (1966). The Advanced Theory of Statistics, Vol. 3. London: Griffin.

Makridakis, S. and Wheelwright, S. (1972). The Box-Jenkins method of forecasting. Europ. Marktng. Rev., 7, 1-18.

Muth, J.F. (1960). Optimal properties of exponentially weighted forecasts. J. Amer. Statist. Assoc., 55, 299-306.

Naylor, T.M. Seaks, T.G. and Wichern, D.W. (1972). Box-Jenkins methods: an alternative to econometric models. Int. Stat. Rev., 40, 123-137.

Newbold, P. (1973 a). The principles of the Box-Jenkins approach. Paper to conference on "Practical Aspects of Forecasting", organized by R. Statist. Soc. Ind. Appl. Section and Oper. Res. Soc. Forecasting Study Group. Imperial College London, April 1972.

Newbold, P. (1973 b). Bayesian estimation of Box-Jenkins transfer function-noise models. J.R. Statist. Soc. B, 35, 323-336.

Newbold, P. and Granger, C.W.J. (1974). Experience with forecasting univariate time-series and the combination of forecasts. J.R. Statist. Soc. A, 137, 131-165.

Reid, D.J. (1969). A comparative study of time-series prediction techniques on economic data. Ph.D. Thesis, University of Nottingham.

Reid, D.J. (1971). A comparison of forecasting techniques on economic time-series. Paper presented at conference organised by Society for Long-range Planning and Forecasting Study Group of Oper. Res. Q.

Thompson, M.E. and Tiao, G.C. (1971). Analysis of telephone data: a case study of forecasting seasonal time-series. Bell J. Econ. Manag. Sci., 2, 514-541.

Tomasek, O. (1972). Statistical forecasting of telephone time-series. I.T.O. Telecommunication Journal, Geneva, December 1972, 1-7.

Walker, A.M. (1967). Some tests of separate families of hypotheses in time-series analysis. Biometrika, 54, 39-68.

Whittle, P. (1952). Tests of fit in time-series. Biometrika, 39, 309-318.

Wilson, G.T. (1973). Contribution to discussion of "Box-Jenkins seasonal forecasting: problems in a case study", by C. Chatfield and D.L. Prothero. J.R. Statist. Soc. A, 136, 315-319.

Winters, P.R. (1960). Forecasting sales by exponentially weighted moving averages. Man. Sci., 6, 324-342.

APPENDIX I

SERIES I

1	2	3	4	5	6	7	8	9	10	11	12	13
153	189	221	215	302	223	201	173	121	106	86	87	108
14	15	16	17	18	19	20	21	22	23	24	25	26
133	177	241	228	283	255	238	164	128	108	87	74	95
27	28	29	30	31	32	33	34	35	36	37	38	39
145	200	187	201	292	220	233	172	119	81	65	76	74
40	41	42	43	44	45	46	47	48	49	50	51	52
111	170	243	178	248	202	163	139	120	96	95	53	94
53	54	55	56	57								
104	135	224	203	210								

SERIES II

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
1965	695	684	817	800	773	807	712	610	499	842	801	722
1966	684	668	854	765	692	751	635	608	501	794	746	678
1967	564	509	670	710	745	780	627	517	547	665	618	615
1968	630	624	767	729	811	781	737	635	563	885	785	679
1969	645	662	722	754	795	798	662	555	709	817	706	639
1970	539	598	646	691	699	800	641	526	489	630	436	425
1971	586	637	756	737	748	798	668	566	756	934	848	649

SERIES III

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
1961	75	65	85	230	385	120	110	115	160	130	110	80
1962	85	75	105	270	430	135	115	130	160	155	120	85
1963	85	80	95	270	445	135	135	145	165	160	135	95
1964	85	90	110	280	465	130	120	120	170	145	120	95
1965	85	90	115	315	520	150	155	145	190	175	150	125
1966	95	105	120	345	590	165	155	165	200	190	155	130
1967	110	95	125	380	615	175	185	200	215	200	170	135
1968	120	115	125	370	605	190	165	170	205	200	160	115
1969	115	110	135	370	590	185	170	170	235	210	155	130

SERIES IV

	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1-4	57461	52558	49452	48062
5-8	53440	46963	48672	50454
9-12	53194	51547	54812	58626
13-16	62814	62073	59406	54932
17-20	59253	55530	55456	57343
21-24	64193	57327	55984	52816
25-28	62657	56583	54271	57484
29-32	68617	61148	60504	59856
33-36	70051	57702	56861	55181
37-40	63931	57200	55550	58300
41-44	65552	62428	59217	57467
45-48	70194	57111	57924	53911