PART VI

SEASONAL MODELS

At a given location on the earth's surface, **hydrologic phenomena,** as well as other natural occurrences tend to vary from one season to another. In many cases, the change in behaviour of a specified phenomenon can be quite spectacular. For example, in Northern Canada the winters are extremely cold whereas the summer temperatures are quite moderate. In addition, due to the melting of snow during the early springtime, the spring flows of northern rivers tend to be quite high while the winter flows are much less. The foregoing and other kinds of hydrologic changes across seasons, are largely dependent upon the rotation of the earth about the sun along with the accompanying changes in tilt of the earth's axis. Local conditions such as mountain ranges, large bodies of water and continental land masses, also influence the hydrologic characteristics of a region.

Seasonal hydrological data are available as time series for which the time intervals between adjacent observations are less than one year. For instance, natural time series are commonly available as average daily, weekly, and monthly sequences. As an illustration of an average monthly time series consider Figure VI.1 which shows the last ten years of the **average monthly flows** of the Saugeen River at Walkerton, Ontario, which are available from January 1915 until December 1976 (Environment Canada, 1977). The sinusoidal shape of this graph is caused by the seasonal changes of the flows throughout the year. As can be seen, the riverflows are high during the spring months and much lower in the winter time. Because the flows vary across the seasons or periods in a cyclic manner, the time series in Figure VI.1 is often referred to as a periodic series.

Except for some random variation, the flows within a given season in Figure VI.1 tend to be stationary across the years. For example, the individual April flows of the Saugeen River vary about the April mean across all of the years from 1915 to 1976. There appears to be no upward or downward trend across the years in the April flows. However, not all periodic series are stationary within each season. Often series which are dependent upon social and economic behaviour possess statistical characteristics which vary across the years within each season. Consider, for example, the graph in Figure VI.2 of the **average monthly water consumption** in millions of litres per day from 1966 to 1988 for the city of London, Ontario, Canada. These observations are available from the Public Utilities Commission (1989) of London. Like other prosperous cities in North America, the city of London has grown dramatically in size since World War II to a 1991 population of about 300,000 people. This increasing population growth, in conjunction with other socio-economic factors have caused the water demand to increase greatly with time, as portrayed by the obvious upward trend in Figure VI.2. Moreover, the seasonality is clearly visible as a sinusoidal pattern wrapped around the trend.

The **greenhouse effect** is referred to in Section 1.2.1. One of the principal greenhouse gases is carbon dioxide (CO_2) which could in conjunction with other gases lead to serious global warning. Figure VI.3 displays a graph of the average monthly concentrations of atmospheric CO_2 measured at the Mauna Loa Observatory located at 3397 meters above sea level on the shoulder of Mauna Loa on the Island of Hawaii. Because the Mauna Loa Observatory is remote

from industrial and other man-made sources of CO_2, the CO_2 observations collected there are believed to be a reliable indicator of the regional trend in the concentration of atmospheric CO_2 in the middle layers of the troposphere. Moreover, the Mauna Loa CO_2 program constitutes the first continuous station anywhere to produce reliable CO_2 data which are listed by both Keeling et al. (1982) and Bacastow and Keeling (1981). As depicted in Figure VI.3 there is a distinct upward trend in the CO_2 data which is probably caused by man-induced interventions such as the large-scale burning of fossil fuels and the massive destruction of forests throughout the world. Consequently, this series provides a vivid illustration of how a natural phenomena can be significantly altered by civilization. One can also see in Figure VI.3 the periodic nature of the CO_2 data which fluctuate in a sinusoidal manner about the trend.

Within section VI, the following three families of seasonal models are presented for modelling seasonal data:

1. **seasonal autoregressive integrated moving average (SARIMA) models** (Chapter 12);
2. **deseasonalized models** (Chapter 13);
3. **periodic models** (Chapter 14).

The deseasonalized and periodic models are used for describing data such as the average monthly flows shown in Figure VI.1, which possess stationarity within each season. The SARIMA family of models can be fitted to data such as those shown in Figures VI.2 and VI.3 where the level and perhaps other statistics change within each season across the years.

The **designs** of all three classes of seasonal models constitute direct extensions of nonseasonal models to the seasonal or periodic cases. More specifically, the SARIMA family of models is a seasonal version of the ARIMA class of models described in Chapters 3 and 4 of Part II. For a given series such as the one in Figure VI.2 or VI.3, nonseasonal and seasonal nonstationarities are removed using nonseasonal and seasonal differencing operators, respectively, before fitting a stationary seasonal ARMA model to the series. To fit a deseasonalized model to a periodic series like the one in Figure VI.1, each observation in the series is first deseasonalized by subtracting out the seasonal mean and then dividing this by the seasonal standard deviation. A nonseasonal ARMA model from Chapter 3 can then be fitted to the resulting nonseasonal series. Finally, a periodic model is formed by fitting a separate PAR or AR model to each season of the year to form what are called **periodic autoregressive (PAR) models** or **periodic autoregressive-moving average (PARMA) models,** respectively. In this way, the varying correlation structure across the seasons in a series such as in the one shown in Figure VI.1 can be directly modelled.

In the next three chapters, each of the three types of seasonal models is defined and **model construction** procedures are presented. Hydrologic and other kinds of applications are employed for demonstrating how the models can be conveniently applied in practice to seasonal data. Because the deseasonalized and periodic models possess quite a few model parameters, procedures are given for reducing the number of model parameters. **Forecasting experiments** in Chapter 15, demonstrate that a certain type of PAR model forecasts seasonal hydrologic time series better than other kinds of competing seasonal models. Consequently, when sufficient data are available, periodic models are the best type of models to use for describing hydrologic and other kinds of natural time series.

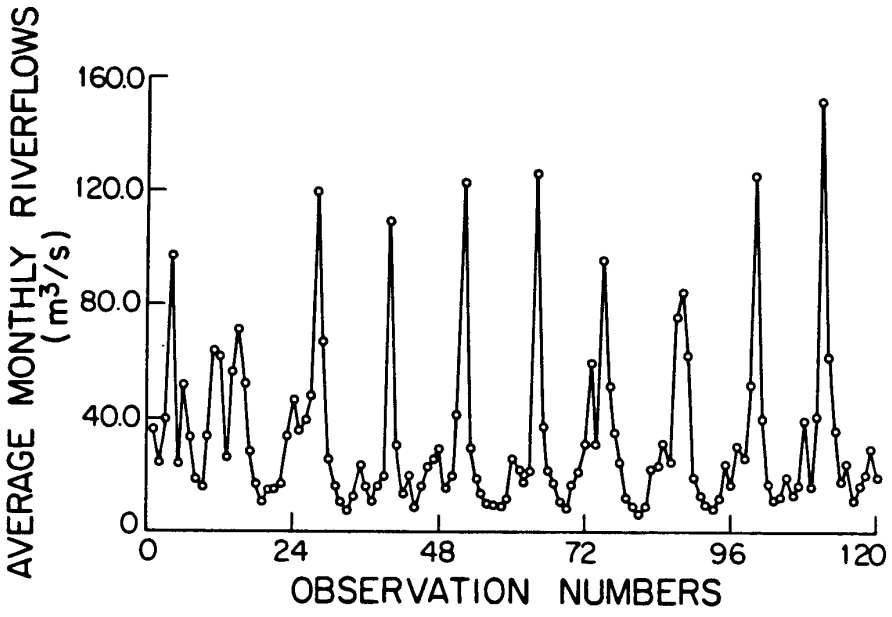

Figure VI.1. Average monthly flows (m^3/s) of the Saugeen River at Walkerton, Ontario, Canada, from January, 1967, until December, 1976.

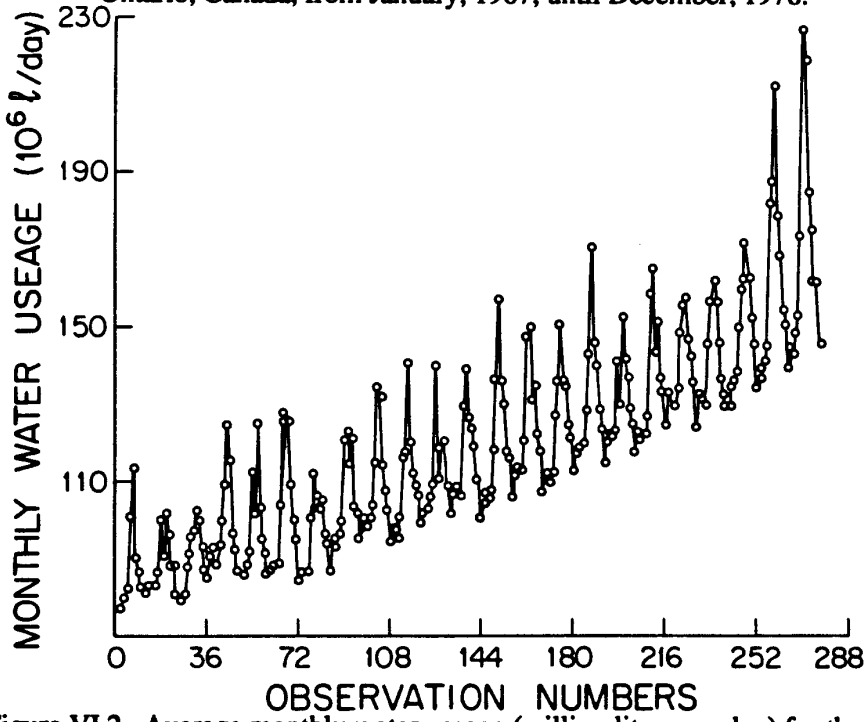

Figure VI.2. Average monthly water useage (million litres per day) for the city of London, Ontario, Canada, from January, 1966, to December, 1988.

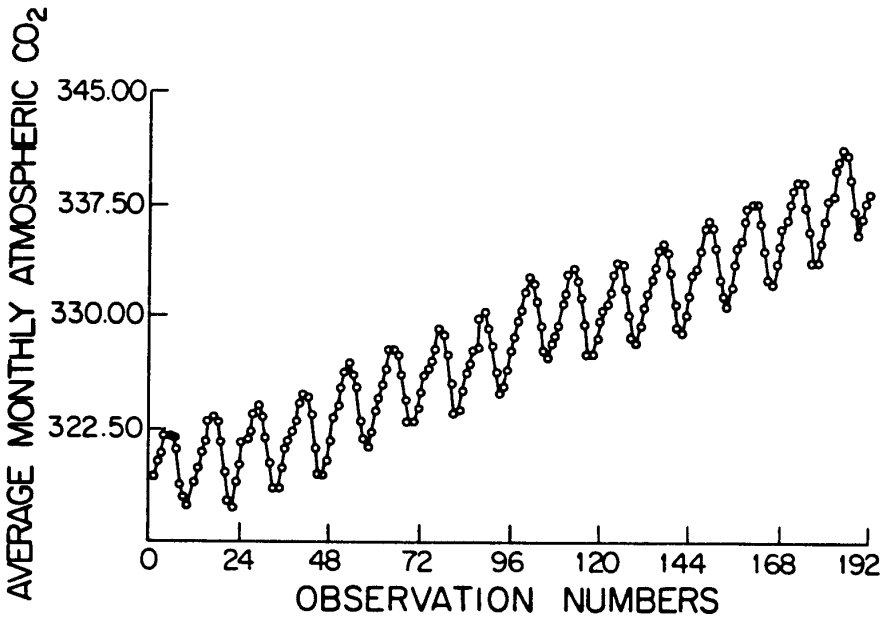

Figure VI.3. Average monthly concentrations of atmospheric CO_2 (molefraction in ppm) measured at Mauna Loa Observatory in Hawaii from January, 1965, to December, 1980.

Although all three seasonal families of models can be used for forecasting hydrologic series, only the deseasonalized and periodic models are properly designed for simulating the seasonally stationary type of data such as the time series shown in Figure VI.1. Additionally, only the SARIMA model contains the appropriate model parameters for describing the nonstationary seasonal data of Figures VI.2 and VI.3. If, for example, the residuals of a SARIMA model fitted to a seasonal series possess correlation which varies from season to season, a PAR model could be fitted to the residuals to capture this behaviour.

CHAPTER 12

SEASONAL

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

MODELS

12.1 INTRODUCTION

Seasonal autoregressive integrated moving average (SARIMA) models are useful for modelling seasonal time series in which the mean and other statistics for a given season are not stationary across the years. The graphs of the average monthly water consumption and atmospheric CO_2 series displayed in Figures VI.2 and VI.3, respectively, depict revealing illustrations of this type of behaviour. Notice that both data sets possess increasing trends which in turn means that the level of each monthly observation is growing in magnitude within the same month over the years. Moreover, the sinusoidal curves that follow the upward trends confirm that the data are seasonal. Figures VI.2 and VI.3 are representative of the general types of nonstationary statistical characteristics that are often present in many kinds of socio-economic time series and natural time series that are significantly affected by man-induced changes, respectively. Other examples of time series which would behave in a similar fashion to the one in Figures VI.2 and VI.3 include average monthly irrigation water consumption, average weekly electricity demand and total quarterly income for recreational facilities located near lakes and rivers.

In the next section, the *mathematical design* of the SARIMA model is presented and associated theoretical properties are described. An inspection of this design indicates why the SARIMA model is ideally suited for modelling a seasonal nonstationary time series like those shown in Figures VI.2 and VI.3 using relatively few model parameters. However, because the mathematical definition does not contain model parameters which explicitly account for separate means and variances in each season, the SARIMA model is not suitably designed for describing series having stationarity of second order moments within each season across the years. For instance, because the average monthly riverflow series of the Saugeen River at Walkerton, Ontario, Canada, plotted in Figure VI.1, appears to have a seasonal mean and variance which are more or less stationary across all the years for each season, a SARIMA is not the best model to fit to this series. Rather, the *deseasonalized and periodic models* of Chapters 13 and 14, respectively, can be employed for modelling this series. Nonetheless, when one is confronted with modelling a seasonal series similar to the one in Figure VI.1, the model construction techniques of Section 12.3 can be utilized for conveniently fitting an appropriate SARIMA model to the series. The *applications* contained in Section 12.4 clearly explain how SARIMA models are fitted in practice by following the identification, estimation and diagnostic check stages of *model construction*. Subsequent to fitting the most appropriate SARIMA model to a series, the calibrated model can be used for purposes such as *forecasting and simulation*, as explained in Section 12.5.

12.2 MODEL DESIGN

12.2.1 Definition

The SARIMA model defined in this section constitutes a straightforward extension of the nonseasonal ARMA and ARIMA models presented in Chapters 3 and 4, respectively. In their book, Box and Jenkins (1976, Ch. 9) define this model and justify why it is useful for describing certain kinds of seasonal series.

Let $z_1, z_2, \ldots, z_\eta$, represent a sequence of seasonal observations. If, for example, there were n years of data for which each year contains s seasons, this would mean that η is equal to ns. For the case of 15 years of monthly data, there would be a total of $15 \times 12 = 180$ observations. If the seasonal time series were not normally distributed and/or the variance of the series changes over time (i.e., the series is heteroscedastic), one could alleviate this problem by invoking a *Box-Cox transformation* (Box and Cox, 1964) defined in [3.4.30] as

$$z_t^{(\lambda)} = \begin{cases} \lambda^{-1}[(z_t + c)^\lambda - 1], & \lambda \neq 0 \\ \ln(z_t + c), & \lambda = 0 \end{cases} \qquad [12.2.1]$$

The parameter λ is the Box-Cox power transformation and c is a positive number which is chosen to be just large enough to cause all the entries in the time series to be positive. If nonnormality and heteroscedasticity in the given series were not detected prior to fitting a SARIMA model to the data, these problems would show up in the residuals of the fitted model. At that time, an appropriate Box-Cox transformation could be selected and the parameters of the SARIMA model could then be estimated for the transformed series.

Figures VI.2 and VI.3 graphically depicts how the magnitudes of observations can change across the seasons in a cyclic manner and also from year to year within a given season. To eliminate nonstationarity within each season, one can employ the *seasonal differencing operator* defined by

$$\nabla_s z_t^{(\lambda)} = (1 - B^s) z_t^{(\lambda)} = z_t^{(\lambda)} - z_{t-s}^{(\lambda)} \quad \text{for } t = s+1, s+2, \cdots, \eta \qquad [12.2.2]$$

where s is the number of seasons per year and B^s is the backward shift operator defined by $B^s z_t^{(\lambda)} = z_{t-s}^{(\lambda)}$. When dealing with monthly data, notice that the relationship in [12.2.2] only connects observations within the same season. Hence, when using seasonal differencing with monthly data, an observation in March is only subtracted from the observation in March of the previous year. If the $z_t^{(\lambda)}$ series is of length $\eta = sn$, the number of observations in the differenced series is $\eta - s$. The differencing operator in [12.2.2] is applied just enough times to remove the seasonal nonstationarity. If it were necessary to apply the seasonal differencing operator in [12.2.2] D times to produce a series of length $\eta - sD$, the resulting series would be given by

$$\nabla_s^D z_t^{(\lambda)} = (1 - B^s)^D z_t^{(\lambda)} \qquad [12.2.3]$$

For purposes of explanation, consider once again a time series consisting of monthly observations. To model correlation among, say, March observations in the differenced series, one may wish to introduce appropriate model parameters. More specifically, to accomplish this task of linking March observations together one can use a model of the form

$$\Phi(B^s)\nabla_s^D z_t^{(\lambda)} = \Theta(B^s)\alpha_t \qquad [12.2.4]$$

where $\Phi(B^s)$ and $\Theta(B^s)$ are the seasonal autoregressive (AR) and seasonal moving average (MA) operators, respectively, and α_t is a residual series which may contain nonseasonal correlation. Both the seasonal AR and MA operators are defined in order to describe relationships within the same season. In particular, the *seasonal AR operator* is defined as

$$\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$$

where Φ_i is the ith AR parameter and P is the order of the AR operator. Because the power of each differencing operator is always an integer multiple of s, only the observations within each season are related to one another when using this operator. Hence, for the case of March observations in a monthly series, only the March observations are connected together using $\Phi(B^s)$. To describe the relationship of the residuals, α_t, within a given season, the *seasonal MA operator* is defined using

$$\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}$$

where Θ_i is the ith MA parameter and Q is the order of the MA operator. Since the exponents of B in $\Theta(B^s)$ are always integer multiples of s, the residuals in the same season are linked with another when using $\Theta(B^s)$.

Theoretically one could define a separate model as in [12.2.4] for each season of the year. However, to keep the model as parsimonious as possible, one can assume that [12.2.4] can be used for all of the seasons. Therefore, one is making the assumption that the correlation within all of the seasons is the same. For the case of monthly data this means that the relationship among all of the March observations is exactly the same as each of the other months.

The error components or residuals, α_t, may contain nonseasonal nonstationarity which can be removed by using the *nonseasonal differencing operator* defined in [4.3.3] as

$$\nabla^d \alpha_t = (1 - B)^d \alpha_t \qquad [12.2.5]$$

where d is the order of the nonseasonal differencing operator which is selected just large enough to remove all of the nonseasonal nonstationarity. The sequence produced using [12.2.5] is theoretically a stationary nonseasonal series. The nonseasonal correlation can then be captured by writing the ARMA model in [3.4.4] or [4.3.4] as

$$\phi(B)\nabla^d \alpha_t = \theta(B)a_t \qquad [12.2.6]$$

where $\phi(B)$ is the *nonseasonal AR operator* of order p defined as

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

and $\theta(B)$ is the *nonseasonal MA operator* of order q written as

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

The a_t's are the innovations which are identically independently distributed (IID) with a mean of zero and variance of σ_a^2. Hence $a_t \sim IID(0,\sigma_a^2)$. In order to obtain maximum likelihood estimates for the model parameters of a SARIMA model, in the next section the restriction of normality is

also placed upon the a_t's so that they are assumed to be distributed as $NID(0,\sigma_a^2)$.

Because of the form of [12.2.6], the correlation among seasons is the same no matter what season one is dealing with. Hence, when entertaining monthly data, the correlation between, say, the March and February observations is defined to be the same as that between any other adjacent months such as October and September. For the periodic models of Chapter 14, this restriction is dropped by allowing for a separate correlation structure for each season of the year.

To define the overall seasonal model, one simply combines equations [12.2.6] and [12.2.4]. This can be accomplished by solving for α_t in [12.2.6] and substituting this result into [12.2.4] to obtain the *SARIMA (seasonal autoregressive integrated moving average) model*

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D z_t^{(\lambda)} = \theta(B)\Theta(B^s)a_t \qquad [12.2.7]$$

Because the operators in [12.2.7] are multiplied together rather than summed, this model is often called a multiplicative SARIMA model.

When fitting the SARIMA model to a given time series of length η, one first transforms the data, if necessary, using the Box-Cox transformation in [12.2.1]. Following this, the data can be differenced both seasonally and nonseasonally. It does not matter which differencing operation is carried out first. One then obtains the stationary series given by

$$w_t = \nabla^d\nabla_s^D z_t^{(\lambda)} \qquad [12.2.8]$$

where the length of the w_t series is $\eta' = \eta - d - sD$. The seasonal and nonseasonal correlation in the w_t sequence is modelled by using the seasonal and nonseasonal AR and MA operators, respectively. Hence, w_t is modelled by employing

$$\phi(B)\Phi(B^s)w_t = \theta(B)\Theta(B^s)a_t \qquad [12.2.9]$$

In some applications, w_t may be a stationary seasonal series which is not obtained by differencing the original series. The model in [12.2.9] is called a *seasonal ARMA or SARMA model* of the w_t series.

12.2.2 Notation

For a given application, one may first wish to indicate the key parameters included in a SARIMA model, without writing down all of the parameter estimates either in a table or else using the difference equation in [12.2.7]. An economical notation for summarizing the structure of the SARIMA model in [12.2.7] is $(p,d,q)\times(P,D,Q)_s$. The first set of brackets contains the orders of the nonseasonal operators while the orders of the seasonal operators are listed inside the second set of brackets. More specifically, p, d and q stand for the orders of the nonseasonal AR, differencing and MA operators, respectively. In the second set of brackets, P, D and Q give the orders of the seasonal AR, differencing and MA operators, respectively. The subscript s appearing to the right of the second set of brackets points out the number of seasons per year.

For the case of monthly data for which $s = 12$, a specific example of a SARIMA model is $(2,1,1)\times(1,1,2)_{12}$. Suppose that the original series were transformed using natural logarithms. By utilizing [12.2.7], this model is written using a finite difference equation as

$$(1-\phi_1 B-\phi_2 B)(1-\Phi_1 B^{12})(1-B)(1-B^{12})\ln(z_t)=(1-\theta_1 B)(1-\Theta_1 B^{12}-\Theta_2 B^{24})a_t$$

If the data are stationary, nonseasonal or seasonal differencing is not required. A stationary model is indicated as $(p,0,q)\times(P,0,Q)_s$. Because this model is stationary, sometimes it is referred to as a SARMA (i.e., seasonal autoregressive-moving average) $(p,q)\times(P,Q)_s$ model.

The summary notation for a pure MA model is $(0,d,q)\times(0,D,Q)_s$. When a model contains no MA parameters, the SARIMA model is written as $(p,d,0)\times(P,D,0)_s$.

When a model is purely nonseasonal, the notation of Part II should be used. Hence, a stationary nonseasonal ARMA model is simply indicated by ARMA(p,q) instead of SARMA(p,q)$\times(0,0)_1$. Likewise, a nonstationary nonseasonal ARIMA model is denoted as ARIMA(p,d,q) rather than the more cumbersome notation given by SARIMA(p,d,q)$\times(0,0,0)_1$.

12.2.3 Stationarity and Invertibility

For a nonseasonal model, the conditions of stationarity and invertibility are discussed in detail in Sections 3.2.2, 3.3.2, and 3.4.2. Recall that for an ARMA model to be stationary the roots of the characteristics equation $\phi(B)=0$ must lie outside the unit circle. Likewise, for invertibility the roots of $\theta(B)=0$ must fall outside the unit circle.

In addition to the aforesaid conditions, the properties of the seasonal AR and MA operators must be specified for the SARIMA model if it is to be fitted to the stationary w_t series in [12.2.9]. For *seasonal stationarity* the roots of the characteristic equation $\Phi(B^s)=0$ must lie outside the unit circle. Similarly, for *seasonal invertibility*, the roots of the characteristic equation $\Theta(B^s)=0$ must fall outside the unit circle.

12.2.4 Unfactored and Nonmultiplicative Models

The SARIMA(p,d,q)$\times$(P,D,Q)$_s$ model in [12.2.7] is referred to as a *multiplicative model.* This is because the nonseasonal and seasonal AR operators are multiplied together on the left hand side while the two MA operators are multiplied together on the right hand side. In addition, the nonseasonal and seasonal differencing operators are multiplied together with the AR operators in [12.2.7]. Following similar arguments, the SARMA(p,q)$\times$(P,Q)$_s$ model which can be fitted to the stationary w_t series in [12.2.9] is also multiplicative.

Rather than write the SARIMA or SARMA model in multiplicative form, it is sometimes useful to use other formats. One approach is to write the models in unfactored form. This is accomplished by multiplying the $\phi(B)$ and $\Phi(B^s)$ operators together to create a single AR operator which can be labelled as $\tilde{\phi}(B)$. Likewise, one can multiply $\theta(B)$ and $\Theta(B^s)$ together to form the single MA operator $\tilde{\theta}(B)$. The SARIMA model in [12.2.7] can then be written in *unfactored form* as

$$\tilde{\phi}(B)\nabla^d\nabla_s^D z_t^{(\lambda)}=\tilde{\theta}(B)a_t \qquad [12.2.10]$$

while the unfactored SARMA model for the w_t series is

$$\tilde{\phi}(B)w_t^{(\lambda)} = \tilde{\theta}(B)a_t \qquad [12.2.11]$$

As mentioned in Section 12.5, the unfactored form of the model in [12.2.10] or [12.2.11] is used for simulating data when differencing operators are present.

As an example of how an unfactored model is determined, consider a SARMA(2,1)×(1,2)$_s$ model which is written in multiplicative form as

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \Phi_1 B^s)w_t = (1 - \theta_1 B)(1 - \Theta_1 B^s - \Theta_2 B^{2s})a_t$$

The unfactored version of this model is

$$(1 - \phi_1 B - \phi_2 B^2 - \Phi_1 B^s + \phi_1\Phi_1 B^{s+1} + \phi_2\Phi_1 B^{s+2})w_t$$
$$= (1 - \theta_1 B - \Theta_1 B^s + \theta_1\Theta_1 B^{s+1} - \Theta_2 B^{2s} + \theta_1\Theta_2 B^{2s+1})a_t$$

In the unfactored model in [12.2.10], the nonseasonal and seasonal differencing operators are not included in the unfactored AR operator $\tilde{\phi}(B)$. To obtain a SARIMA model that does not have separate differencing factors, one can write the model in *generalized form* as

$$\phi'(B)z_t^{(\lambda)} = \theta'(B)a_t \qquad [12.2.12]$$

where $\phi'(B) = \phi(B)\Phi(B^s)\nabla^d\nabla_s^D$ is the *generalized or unfactored AR operator* for which ϕ'_i is the ith generalized AR parameter, and $\theta'(B) = \theta(B)\theta(B^s) = \tilde{\theta}(B)$ is the *generalized or unfactored MA operator* for which θ'_i is the ith generalized MA parameter. The unfactored format in [12.2.12] is useful for calculating minimum mean squared error forecasts, as is pointed out in Section 12.5.

Because the unfactored models are equivalent to their multiplicative counterparts, unfactored models are, of course, multiplicative. For some applications one may wish to generalize the multiplicative models given in equations [12.2.7], [12.2.9], [12.2.10] and [12.2.11], by allowing nonmultiplicative AR and MA operators, which are denoted by $\phi^*(B)$ and $\theta^*(B)$, respectively. If one wishes to remove nonseasonal and seasonal stationarity using differencing, the *nonmultiplicative model* is given as

$$\phi^*(B)\nabla^d\nabla_s^D z_t^{(\lambda)} = \theta^*(B)a_t \qquad [12.2.13]$$

When the given series is already stationary due to differencing or some other type of operation, the nonmultiplicative model is written as

$$\phi^*(B)w_t = \theta^*(B)a_t \qquad [12.2.14]$$

The model in [12.2.14] is, in fact, the same as the nonseasonal ARMA model given in [3.4.4]. However, in practice one would not use all of the parameters in the ARMA model and, therefore, use the constrained type of model discussed in Section 3.4.4. An example of a *constrained model* is

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_s B^s - \phi_{s+1} B^{s+1})w_t = (1 - \theta_1 B - \theta_s B^s - \theta_{s+1} B^{s+1})a_t$$

Notice that this model cannot be factored to form a multiplicative model, as is the case for the earlier example for the unfactored model.

In addition to entertaining the nonmultiplicative model, certain problems may dictate that a more complex multiplicative seasonal model should be used. For example, suppose that a water demand series contains two distinct seasonal components. One component may be due to industrial use which follows a different seasonal pattern than the second component, which is the residential demand. For each seasonal component, separate AR and MA operators could be defined and all of these operators would be combined together in an overall multiplicative seasonal model.

12.2.5 Autocorrelation Function

The derivations of the theoretical ACF for nonseasonal AR, MA, and ARMA processes are presented in Sections 3.2.2, 3.3.2 and 3.4.2, respectively. One could follow a similar approach to that used for the nonseasonal case to develop the formula for the theoretical ACF of the stationary seasonal process, w_t, given in [12.2.9]. However, a simpler approach is to write the SARMA model for w_t in [12.2.9] in the same format as its nonseasonal counterpart by using the unfactored form in [12.2.11]. Subsequent to this, the procedure developed for the nonseasonal case can be used to obtain the theoretical ACF for the seasonal model. To use the theoretical results for the ACF developed for the nonseasonal ARMA model in Section 3.4.2, simply replace the nonseasonal AR and MA operators by their combined counterparts for the seasonal model. The algorithm of McLeod (1975) presented in Appendix A3.2 can then be used to determine the theoretical ACF for the SARMA process. In a similar fashion, one can also calculate the theoretical PACF for the seasonal model by utilizing appropriate results developed for the nonseasonal case in Chapter 3.

12.2.6 Three Formulations of the Seasonal Processes

Introduction

The three formulations for the ARMA and ARIMA processes are given in Sections 3.4.3 and 4.3.4, respectively. As explained in these sections, the ARMA and ARIMA processes can be written using the original forms for their finite difference equations, the purely MA or random shock formulations, or the purely AR or inverted formats. Likewise, one can write the SARIMA or SARMA processes using any of the three formulations. The difference equation forms in which the SARIMA and SARMA processes are originally defined are given in [12.2.7] and [12.2.9], respectively. The random shock and inverted formulations are now described.

Random Shock Form

Because of both the invertibility and stationarity conditions referred to in Section 12.2.3, one can manipulate the operators algebraically and, hence, write the SARIMA process as either a pure MA process or a pure AR process. Using [12.2.7], the pure MA or *random shock form of the SARIMA model* is

$$z_t^{(\lambda)} = \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)\nabla^d\nabla_s^D} a_t$$

$$= a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \cdots$$
$$= a_t + \Psi_1 B a_t + \Psi_2 B^2 a_t + \cdots$$
$$= (1 + \Psi_1 B + \Psi_2 B^2 + \cdots) a_t$$
$$= \Psi(B) a_t \qquad [12.2.15]$$

where $\Psi(B) = 1 + \Psi_1 B + \Psi_2 B^2 + \cdots$, is the *seasonal random shock or infinite MA operator,* and Ψ_i is the ith parameter, coefficient or weight of $\Psi(B)$.

It is instructive to express a SARIMA process in the random shock in [12.2.15] for both theoretical and practical reasons. For example, the Ψ weights are needed to calculate the variance of forecasts when using a SARIMA model for forecasting. Also, one way to simulate data is to first write a SARIMA model in random shock form and then use this format of the model for producing synthetic data. Finally, by writing each member of a set of SARIMA models in random shock form, the models can be conveniently compared by examining the magnitudes and signs of the Ψ parameters.

In practice, one would first fit a SARIMA model to a given time series by following the model building approach described in Section 12.3 in order to obtain estimates for the AR and MA parameters. Subsequent to this, one may wish to calculate the Ψ_i parameters in [12.2.15], given the AR and MA parameters. The relationship for carrying out these calculations is contained in [12.2.15]. By definition, the following identity is true:

$$\Psi(B) = \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)\nabla^d \nabla_s^D}$$

or

$$\phi(B)\Phi(B^s)\nabla^d \nabla_s^D \Psi(B) = \theta(B)\Theta(B^s) \qquad [12.2.16]$$

By equating coefficients of B^k, $k = 1,2,\cdots$, on the left hand side of [12.2.16] to those on the right hand side, one can use the identity to solve for the Ψ_i coefficients in terms of the AR and MA parameters. Examples of performing these manipulations for nonseasonal ARMA models are presented in Section 3.4.3. Similar calculations can be carried out using [12.2.16] or a simplified version thereof for the seasonal case.

If no differencing operators are present and one is dealing with the SARMA model in [12.2.9] instead of the SARIMA model in [12.2.7], a similar procedure can be used to calculate the random shock weights. Simply remove the differencing operators from [12.2.16] and, once again, compare coefficients of B^k, $k = 0,1,2,\cdots$, to solve for the random shock parameters.

Another approach to solve for the random shock parameters for the SARIMA or SARMA model is to use [4.3.11] or [3.4.21] to solve for the Ψ_i's. This is accomplished by writing the SARIMA or SARMA model in unfactored form by using [12.2.12] or [12.2.11], respectively, and then making appropriate substitutions for the operators so that the nonseasonal formulae for calculating the seasonal random shock parameters can be used.

Inverted Form

To express the SARIMA model in inverted form, equation [12.2.7] can be written as

$$\begin{aligned} a_t &= \frac{\phi(B)\Phi(B^s)\nabla^d\nabla_s^D z_t^{(\lambda)}}{\theta(B)\Theta(B^s)} \\ &= z_t^{(\lambda)} - \Pi_1 z_{t-1}^{(\lambda)} - \Pi_2 z_{t-2}^{(\lambda)} - \cdots \\ &= z_t^{(\lambda)} - \Pi_1 B z_t^{(\lambda)} - \Pi_2 B^2 z_t^{(\lambda)} - \cdots \\ &= (1 - \Pi_1 B - \Pi_2 B^2 - \cdots) z_t^{(\lambda)} \\ &= \Pi(B) z_t^{(\lambda)} \end{aligned} \qquad [12.2.17]$$

where $\Pi(B) = 1 - \Pi_1 B - \Pi_2 B^2 - \cdots$, is the *seasonal inverted or infinite AR operator* and Π_i is the ith parameter, coefficient or weight of $\Pi(B)$. By comparing [12.2.17] and [12.2.16], one can see that

$$\Psi(B)^{-1} = \Pi(B) \qquad [12.2.18]$$

To calculate the inverted parameters, given that one knows the values of the AR and MA parameters, one can use the identity

$$\Pi(B) = \frac{\phi(B)\Phi(B^s)\nabla^d\nabla_s^D}{\theta(B)\Theta(B^s)}$$

or

$$\theta(B)\Theta(B^s)\Pi(B) = \phi(B)\Phi(B^s)\nabla^d\nabla_s^D \qquad [12.2.19]$$

which is obtained from [12.2.17]. To solve for the Π_i parameters, simply equate the coefficients of B^k for $k = 0,1,2,\cdots$, on the left hand side of [12.2.16] to those on the right hand side.

If one wishes to calculate the inverted parameters for the SARMA model in [12.2.9], simply eliminate the differencing operators in [12.2.19]. Following this, one can use [12.2.19] to determine the inverted parameters by equating coefficients of the B^k for $k = 0,1,2,\cdots$, on the left hand side to those on the right.

As was done for the random shock parameters, one can also use the nonseasonal formulae to determine the seasonal inverted parameters. Simply write the SARIMA or SARMA model in unfactored form as in [12.2.12] or [12.2.11], respectively, and then make appropriate substitutions into [4.3.14] or [3.4.27], respectively to solve for the seasonal inverted parameters.

12.3 MODEL CONSTRUCTION

12.3.1 Introduction

The most appropriate SARIMA model to fit to a given seasonal time series can be ascertained by following the identification, estimation, and diagnostic check stages of model construction. In the previous section, it is shown how the design of the multiplicative SARIMA family of models is a straightforward extension of the nonseasonal models presented in Part II of the book. Likewise, as is explained in this section, the tools used for SARIMA model building

are either the same or else closely related versions of the nonseasonal model construction techniques discussed in Part III. The applications described in Section 12.4 clearly demonstrate that the model development techniques of this section can be conveniently and easily utilized for determining the most appropriate SARIMA or SARMA model to fit to a given time series. Finally, to allow the model building stages to be expeditiously and properly implemented in practice, one can employ a flexible decision support system such as the McLeod-Hipel Package described in Section 1.7.

12.3.2 Identification

Introduction

The purpose of the identification stage is to determine the nonseasonal and seasonal differencing required to produce stationarity and also the orders of both the nonseasonal and the seasonal AR and MA operators for the w_t series in [12.2.9]. Although each identification technique is discussed separately, in practical applications the output from all the techniques is interpreted and compared together in order to design the type of model to be estimated.

For some types of seasonal time series, it is known in advance whether or not the data sets should be transformed using the Box-Cox transformation in [12.2.1]. For instance, average monthly riverflow series often require a natural logarithmic transformation to cause the residuals of the fitted models to be approximately normally distributed and homoscedastic. In many applications, analysts may not realize that Box-Cox transformations are needed until after the model parameters have been estimated and the statistical properties of the residuals are examined. The analysts should keep in mind that usually a Box-Cox transformation does not change the design of the AR and MA operators needed in the model or models to fit to the transformed time series. However, this is not true in general, and as is pointed out by Granger and Newbold [1976], certain transformations can change the type of model to estimate for a given time series. Therefore, even though it is usually not necessary to perform the identification stage for the transformed data if it has already been done for the corresponding untransformed series, a practitioner should be aware that in some instances this may not be the case. When a transformation does change the type of model to be used, diagnostic checks would detect this fact and then the design of the model to fit to the transformed data can be properly identified.

For a SARIMA model application there should be at least seven years of seasonal data and also at least 50 data points overall in order to get reasonable MLE's (maximum likelihood estimates) for the model parameters. If one were analyzing a monthly series, one would require at least $12 \times 7 = 84$ observations. Therefore, one should proceed with the identification stage only if the minimum required amount of information is present.

Tools

When examining a specified time series analysis for the first time, one may wish to utilize the exploratory data analysis tools described in Section 22.3. The purpose of *exploratory data analysis* is to discover the basic statistical characteristics of a data set by examining simple graphical and numerical output. Subsequent to obtaining a general understanding of the statistical properties of the time series, one may wish to design a specific SARIMA model to fit to the series by studying the graphical output from the following techniques which are also used for designing nonseasonal models in Chapter 5:

1. **Plot of the Original Series** - A graph of the observations in the series against time is an important exploratory data analysis method that should *always* be used in model identification. Characteristics of the data which are usually easily uncovered from a perusal of a time series plot include seasonality, nonstationarity due to trends in the mean levels of the seasons or years, changing variance, extreme values, correlation or dependence among observations, and long term cycles. The nonstationarity present in the data can often be removed using seasonal and/or nonseasonal differencing. The seasonal and nonseasonal correlation in the time series can be modelled by appropriately deciding upon which AR and MA operators should be included in the SARIMA model. Graphs of the next four functions can be used for specifically designing the components needed in the SARIMA model.

2. **ACF (autocorrelation function)** - The theoretical ACF defined in [2.5.4] measures the amount of linear dependence between observations in a time series that are separated by k time lags. The sample estimate, r_k, for ρ_k, is given in [2.5.9] while approximate variances for r_k are given in [2.5.10] and [2.5.11]. To use the sample ACF in model identification, calculate and then plot r_k up to a maximum lag of roughly $\frac{\eta}{4}$ along with the approximate 95% confidence limits. The graph of the sample ACF and the other three graphs described below should include at least $2s$ or $3s$ lags, where s is the number of seasons per year. In this way, the cyclic behaviour caused by seasonality and any decaying or truncation properties of r_k over k, can be visually detected.

   The first step is to examine a plot of the ACF to detect the presence of nonstationarity in the given series. For seasonal data with the seasonal length equal to s, the ACF often follows a wave pattern with peaks at s, $2s$, $3s$, and other integer multiples of s. As is shown by Box and Jenkins (1976, pp. 174-175), if the estimated ACF at lags that are integer multiples of the seasonal length s do not die out rapidly, this may indicate that seasonal differencing is needed to produce stationarity. Failure of other ACF estimates to damp out may imply that nonseasonal differencing is also required. If the length of the original series is η, the number of data points in the differenced series would be $\eta' = \eta - d + sD$. Li (1991) develops some statistical tests for determining the orders of differencing required for a seasonal time series.

   If the stationary w_t series is not white noise, one can use the sample ACF to help decide upon which AR and MA parameters are needed in the SARIMA model. When the process is a pure MA$(0,d,q)\times(0,D,Q)_s$ model, the sample ACF truncates and is not significantly different from zero after lag $q + sQ$. For this case, the variance of r_k after lag $q + sQ$ is (Bartlett, 1946)

$$Var[r_k] \approx \frac{1}{\eta'}\left[1 + 2\sum_{i=1}^{q+sQ} r_i^2\right], \quad k > q + sQ \qquad [12.3.1]$$

   where η' stands for the length of the w_t series after differencing.

   If r_k attenuates at lags that are multiples of s, this implies the presence of a seasonal AR component. The failure of the ACF to truncate at other lags may imply that a nonseasonal AR term is required.

As defined in [12.2.8], the stationary w_t series created by differencing either the original or transformed series is given as

$$w_t = \nabla^d \nabla_s^D z_t^{(\lambda)}$$

where the exponent λ indicates that the original z_t series may be transformed using the Box-Cox transformations in [12.2.1]. After the data have been differenced just enough times to produce both seasonal and nonseasonal stationarity, then check the ACF of the w_t series to determine the number of AR and MA parameters required in the model. The w_t series is also used at the other steps of the identification procedure. Of course, if no differencing is required, the w_t series is simply the $z_t^{(\lambda)}$ series. As noted earlier, the graph of the sample ACF for w_t should include at least $2s$ or $3s$ lags.

If a series is white noise, then r_k is approximately NID(0,$\frac{1}{\eta'}$). This result allows one to test whether a given series is white noise by checking to see if the ACF estimates are significantly different from zero. Simply plot confidence limits on the ACF diagram and see if a significant number of r_k values fall outside the chosen confidence interval.

3. **PACF (partial autocorrelation function)** - After writing the SARIMA or SARMA model in unfactored form as in [12.2.10] or [12.2.11], respectively, the theoretical PACF is defined for the w_t series in [3.2.17] using the Yule-Walker equations. Following the approach discussed in Section 3.2.2 and Appendix A3.1, the sample PACF can be estimated. For model identification, simply calculate and plot the sample PACF to at least lag $2s$ along with the 95% confidence limits which are calculated using [3.2.18], in which the length of the series is taken to be that of w_t. Employing rules put forward by authors such as Hipel et al. (1977), McLeod et al. (1977), and Hamilton and Watts (1978), the sample PACF can be utilized for deciding upon which AR and MA parameters are needed for properly representing the data.

   When the process is a pure AR$(p,d,0)\times(P,D,0)_s$ model, the sample PACF cuts off and is not significantly different from zero after lag $p + sP$. After lag $p + sP$, the sample PACF is approximately NID(0,$\frac{1}{\eta'}$).

   If the sample PACF damps out at lags that are multiples of s, this suggests the incorporation of a seasonal MA component into the model. The failure of the sample PACF to truncate at other lags may imply that a nonseasonal MA term is required.

4 **IACF (inverse autocorrelation function)** - The theoretical IACF (Cleveland, 1972) is defined in Section 5.3.6 and a method for estimating the sample IACF along with approximate 95% confidence limits is given in the same section. Theoretically, the IACF of the w_t series is defined to be the ACF of the $(q,d,p)\times(Q,D,P)_s$ process that is written as

$$\theta(B)\Theta(B^s)w_t = \phi(B)\Phi(B^s)a_t \qquad [12.3.2]$$

The model in [12.3.2] is called the *dual model* while the SARIMA$(p,d,q)\times(P,D,Q)_s$ model in [12.2.7] or [12.2.9] is referred to as the *primal model* (McLeod, 1984).

As is the case for all four functions discussed under points 2 to 5, the sample IACF is plotted up to a lag of at least $2s$ or $3s$ or not more than $\frac{\eta'}{4}$. If the w_t series is white noise, the sample IACF is approximately NID$(0,\frac{1}{\eta'})$. For the case of white noise, the values of the sample IACF should not fall outside the 95% confidence limits of $\pm\frac{1.96}{\sqrt{\eta'}}$ more than once in twenty lags.

For a pure AR$(p,d,0)\times(P,D,0)_s$ process, the sample IACF truncates and is not significantly different from zero after lag $p + sP$. If the sample IACF damps out but is still significant at lags s, $2s$, $3s$, etc., a seasonal MA component may be needed in the model. An additional nonseasonal MA component will cause the sample IACF to damp out for values between 0 and s, s and $2s$, etc., where decreasing but prominent peaks occur at s, $2s$, $3s$, etc., due to the seasonal MA term.

5. **IPACF (inverse partial autocorrelation function)** - The theoretical IPACF originally defined by Hipel et al. (1977) is presented in Section 5.3.7. The PACF for a SARMA model is by definition the IPACF of the dual model in [12.3.2]. In addition, a method for estimating the IPACF and obtaining approximate 95% confidence limits is explained in Section 5.3.7.

   For model identification, the sample IPACF and its 95% confidence limits are plotted up to a lag of at least $2s$ or $3s$. If the w_t series is white noise, then the values of the sample IPACF should not be significantly different from zero and should fall within the 95% confidence limits.

   For a pure MA$(0,d,q)\times(0,D,Q)_s$ model, the sample IPACF truncates and is not significantly different from zero after lag $q + sQ$. After lag $q + sQ$, the sample IPACF is approximately NID$(0,\frac{1}{\eta'})$. If the sample IPACF attenuates at lags that are multiples of s, this may indicate the presence of a seasonal AR component. When the IPACF fails to cut off at other lags, this implies the need for a nonseasonal AR term.

6. **Cumulative periodogram white noise test** - As was mentioned previously, the sample ACF plot is an accepted means of checking whether the given data are white noise. The sample PACF, IACF, and IPACF can also be employed in this capacity. However, the cumulative periodogram defined in [2.6.2] provides another means of checking for white noise.

   In addition to verifying whether a series is uncorrelated, the cumulative periodogram can also detects certain types of correlation. In particular, it is an effective procedure for finding hidden periodicities.

Summary

A plot of the original data portrays an overall view of how the time series is generally behaving and whether or not differencing is required. However, the sample ACF, PACF, IACF, and IPACF transform the given information into a format whereby it is possible to detect the number of AR and MA terms required in the model. In general, the ACF and the IPACF truncate for pure MA processes, while the PACF and IACF cut off for AR models. For mixed

processes, all four functions attenuate. This behaviour of identification graphs for SARIMA models is summarized in Table 12.3.1.

The ACF and the IPACF possess similar general properties, while the PACF and the IACF have common attributes. However, the four functions are defined differently, and none of them behave exactly in the same fashion. In practice, the authors have found that if the PACF fails to detect a certain property of the time series, then the IACF often may be more sensitive and thereby may clearly display the presence of that property and vice versa. A similar situation exists between the ACF and the IPACF. In actual applications, it is necessary to consider simultaneously the output from all the functions in order to ascertain which model to estimate.

The incorporation of the IACF and the IPACF into the identification stage simplifies and substantiates model design because it is easier and more accurate to determine the proper SARIMA model to estimate. It is recommended that all of the identification plots be programmed for instantaneous display on a computer terminal screen. In this way, the identification stage can usually be completed in just a few minutes. The capability of making an immediate copy of any results portrayed on a screen provides a convenient method of keeping a permanent record. The decision support system for time series modelling described in Section 1.7 can be employed in this manner.

In Appendix A12.1, an alternative procedure for using the ACF to identify the parameters required in a SARMA model is given. This novel approach utilizes the structure of the multiplicative SARMA model by splitting the analysis using the ACF into nonseasonal and seasonal components.

Table 12.3.1. Behaviour of identification functions for SARIMA models.

	TYPES OF MODELS		
FUNCTION	Pure AR $(p,d,O)\times(P,D,O)_s$	Pure MA $(O,d,q)\times(O,D,Q)_s$	Mixed $(p,d,q)\times(P,D,Q)_s$
ACF	Attenuates	Truncates after lag $q + sQ$	Attenuates
PACF	Truncates after lag $p + sP$	Attenuates	Attenuates
IACF	Truncates after lag $p + sP$	Attenuates	Attenuates
IPACF	Attenuates	Truncates after lag $q + sQ$	Attenuates

12.3.3 Estimation

Introduction

Often, identification methods cannot clearly determine which is the single best SARIMA model to fit to the time series under study. Rather, anywhere from one to four model designs may be tentatively identified. At the estimation stage, *MLE*'s can then be obtained for the parameters in each of the models. Subsequently, discrimination methods can be used for selecting the best model from the set of calibrated models. The techniques for choosing the best

model include the AIC discussed in this section as well as Section 6.3 and the diagnostic checks presented in Section 12.3.4. If none of the fitted models adequately describes the data, appropriate design modifications can be made before estimating the parameters of the most recent iteration and repeating the above procedure until a suitable model is found.

Algorithms

Two algorithms are discussed in this section for obtaining approximate MLE's for the parameters of the SARMA model fitted to the w_t series in [12.2.9]. Besides using the basic definition of the SARMA model, both methods are based on the assumption that the innovations are normally independently distributed with a mean of zero and variance of σ_a^2 [i.e. NID(0,σ_a^2)]. The first approach is to use the algorithm developed for the nonseasonal ARMA model while the second one is to employ a more computationally efficient procedure which takes into account the specific mathematical structure of the SARMA model. Because the orders of nonseasonal and seasonal differencing operators required to produce the stationary w_t series in [12.2.9] are not estimated but selected based upon identification results, only the parameters included in the SARMA model for fitting to w_t have to be estimated. Consequently, in this section, parameter estimation is discussed in terms of the SARMA$(p,q)\times(P,Q)_s$ model in [12.2.9] rather than the SARIMA$(p,d,q)\times(P,D,Q)_s$ model in [12.2.7]. Furthermore, since the mean level of w_t is zero due to differencing, the mean level or trend is not included in the SARMA model in [12.2.9]. If it were necessary to estimate a mean or trend this could be accomplished using the estimation methods discussed in this section.

In Chapter 6, the *modified sum of squares algorithm of McLeod (1977)* is suggested for estimating the parameters of the ARMA(p,q) model in [3.4.4]. As explained in that chapter, compared to other competing estimation methods, the modified sum of squares approach is both computationally and statistically efficient. The main steps in the algorithm are described in Appendix A6.1. To use the modified sum of squares method in Chapter 6 to estimate the parameters of the SARMA(p,q) model in [12.2.9], the first step is to write the model in the unfactored form given in [12.2.11]. Because the unfactored model in [12.2.11] can be considered as a special case of the ARMA(p,q) model, the modified sum of squares method can be used to obtain approximate MLE's for the model parameters and residuals.

The first estimation method works well when the number of seasons per year is not more than 12. However, for bimonthly data and weekly series for which $s = 24$ and 52, respectively, the estimation becomes computationally inefficient. To overcome this problem, the *maximum likelihood approach of McLeod and Salas (1983)* can be used. This estimation method, which is based upon the modified sum of squares method of McLeod (1977), is designed according to the multiplicative structure of the AR and MA operators in the SARMA model. It works for yearly ($s = 1$), monthly ($s = 12$), weekly ($s = 52$), daily ($s = 365$) as well as any other types of seasonal series. The main steps in the McLeod-Salas algorithm are described in Appendix A12.2.

The MLE's for a SARMA or other kind of time series model are asymptotically normally distributed. Because a *SE (standard error)* is estimated for each of the estimated parameters using the information matrix, one can check if an estimate is significantly different from zero. If the estimate is significant at the 5% significance level, its absolute magnitude should be larger than 1.96 SE. Usually, it is advisable to drop parameters which are not significantly different from zero from the SARMA model and then to re-estimate the parameters of the simplified

model and then check if this model provides an adequate fit.

Model Discrimination

Subsequent to estimating the model parameters for the models separately fitted to the time series under study, one can calculate the value of the AIC for each model in order to select the model which has the minimum AIC value. This procedure is referred to as *MAICE (minimum AIC estimation)* and is described in detail in Section 6.3. The flow chart in Figure 6.3.1 explains the ways in which MAICE can be used in model construction for application purposes. One approach is to carry out an exhaustive AIC study by fitting a large range of SARMA models to the time series and then picking the one having the minimum AIC. In the second main approach, the identification techniques of Section 12.3.2 are used to select a handful of models for which the parameters and AIC values are estimated. Once again, one selects the model having the minimum AIC value.

The general formula for the AIC is given in [6.3.1] as (Akaike, 1974)

$$AIC = -2\ln(ML) + 2k$$

where ML denotes the maximized value of the likelihood function and k is the number of independently adjusted parameters in the model. Approximate formulae can be devised for determining the AIC for a SARIMA model which contains differencing operators. Because the amount of data has been reduced from a total of η to $\eta' = \eta - d - sD$ points when there is both nonseasonal and seasonal differencing, this will certainly affect the first term on the right-hand side of [6.3.1]. Hence, the AIC for a SARIMA model can be roughly calculated as

$$AIC = \frac{\eta}{\eta'}(-2\ln(ML)) + 2k \qquad [12.3.3]$$

where the maximized log likelihood is obtained by optimizing the log likelihood function defined in [A12.2.1]. The total number of model parameters is $k = p + q + P + Q + 1$, where the unity term allows for the estimate of the variance of the model residuals. Usually, the mean of the differenced series can be assumed to be zero. However, if the mean of the differenced series is also estimated, k must be increased by unity. Also, k is increased by one if $\lambda \neq 1$.

Another alternative for developing an AIC formula for a SARIMA model is to alter both of the terms on the right-hand side of [6.3.1]. As argued by Ozaki (1977), an increase in the number of data points contributes to decreasing the penalty due to the number of parameters. When the data are differenced both nonseasonally and seasonally, the number of data points decreases from η to $\eta' = \eta - d - sD$. This effect can be incorporated into the AIC by writing the formula as

$$AIC = \frac{\eta}{\eta'}(-2\ln(ML) + 2k) \qquad [12.3.4]$$

12.3.4 Diagnostic Checks

Introduction

The three assumptions underlying the innovations, a_t, $t = 1,2,\ldots,\eta'$, of the SARMA model in [12.2.9] are that the disturbances are independent, homoscedastic (i.e. have constant

variance) and normally distributed. To check the foregoing residual stipulations, the estimated innovations, $\hat{a}_t$'s, or model residuals are required. The estimates for the $\hat{a}_t$'s are automatically calculated at the estimation stage along with the MLE's and SE's for the SARMA model parameters (see Appendices A12.2, A6.1 and A6.2).

A data transformation cannot correct dependence of the residuals because the lack of independence indicates that the present model is inadequate. Rather, the identification and estimation stages must be repeated in order to determine a suitable model having different model parameters. If the less important assumptions of homoscedasticity and normality are violated, they can often be corrected by a Box-Cox transformation of the data.

The residual assumptions for the SARMA model are identical to those stipulated for the nonseasonal ARMA model in [3.4.4]. Consequently, all of the diagnostic checks presented for the nonseasonal ARMA model in Chapter 7, can be used with the SARMA model. Some of these diagnostic methods are briefly described below but for detailed accounts of these and other model checking methods, the reader can refer to Chapter 7.

Tests for Whiteness

If a calibrated SARMA model adequately describes a time series, the estimated innovations, $\hat{a}_t$'s, or residuals should be white, due to the independence assumption of the a_t's. To determine whether the residuals are white noise, the best procedure is to examine the *residual autocorrelation function (RACF)*. Because the distribution of the RACF which is shown in the theorem below is now known, sensitive testing techniques are available for checking the independence assumption of a_t.

The theorem for the RACF is developed as follows. The ACF, $r_k(\hat{a})$, of the calculated residuals can be determined by

$$r_k(\hat{a}) = \sum_{t=k+1}^{\eta'} \left[\hat{a}_t \hat{a}_{t-k} / \left(\sum_{i=1}^{\eta'} \hat{a}_i^2 \right) \right] \qquad [12.3.5]$$

Define the vector of the first L values of the RACF as

$$\mathbf{r}(\hat{a}) = [r_1(\hat{a}), r_2(\hat{a}), \ldots, r_L(\hat{a})]' \qquad [12.3.6]$$

Denote by $\Psi_k(\Phi)$ the coefficient of B^k in the Maclaurin series expansion of $[\Phi(B^s)]^{-1}$ in powers of B, and similarly define $\Psi_k(\phi)$, $\Psi_k(\Theta)$, and $\Psi_k(\theta)$. Then it can be proved for large samples (McLeod, 1978) that

$$\mathbf{r}(\hat{a}) \approx N[0, (1/\eta')\mathbf{U}] \qquad [12.3.7]$$

where $\mathbf{U} = \mathbf{1}_L - \mathbf{X}\mathbf{T}^{-1}\mathbf{X}$, $\mathbf{1}_L$ is the identity matrix, $\mathbf{I} \approx \mathbf{X}'\mathbf{X}$ is the large-sample information matrix, and $X = [\Psi_{t-js}(\Phi), \Psi_{t-j}(\phi), \Psi_{t-js}(\Theta), \Psi_{t-j}(\theta)]$ are the i,j entries in the four partitions of the $\mathbf{X}$ matrix. The dimensions of the matrices X, $\Psi_{t-js}(\Phi)$, $\Psi_{t-j}(\phi)$, $\Psi_{t-js}(\Theta)$, and $\Psi_{t-j}(\theta)$ are, respectively, $L \times (P + p + Q + q)$, $L \times P$, $L \times p$, $L \times Q$, and $L \times q$. Previously, Box and Pierce (1970) obtained this result for the nonseasonal AR case, but the theorem listed here is valid for nonseasonal ARMA, SARMA, transfer function-noise, and intervention models.

There are two useful applications of the RACF distribution theorem. A sensitive diagnostic check is to first plot the RACF along with the asymptotic significance intervals for the RACF that are obtained from the diagonal entries of the matrix $\left(\frac{1}{\eta'}\right)$U. If some of the RACF values are significantly different from zero, this may mean that the present model is inadequate. The important values of the RACF to examine are those at the first couple of lags and also at lags that are integer multiples of s for a seasonal model.

If the current model is insufficient, one can use the information from the RACF plot to help design an improved model before returning to the earlier stages of model construction. For example, a significantly large value of the RACF at lag s may indicate that a seasonal MA parameter is needed in the SARMA model. Upon updating the model design, the new model can be calibrated and checked for the presence of any further weaknesses.

A second but less sensitive test is to calculate and to perform a significance test for the modified Portmanteau statistic U_L (Li and McLeod, 1981). If L is large enough so that the weights $\Psi_k(\Phi)$, $\Psi_k(\phi)$, $\Psi_k(\Theta)$, and $\Psi_k(\theta)$ have damped out, then

$$U_L = \eta' \sum_{k=1}^{L} r_k^2(\hat{a}) + \frac{L(L+1)}{2\eta'} \qquad [12.3.8]$$

where L can be given as value of $2s$ to $4s$ such that the maximum value of L is not more than about $\frac{\eta'}{4}$. The statistic U_L is χ^2 distributed on $(L - P - p - Q - q)$ degrees of freedom. A test of this hypothesis can be done for model adequacy by choosing a level of significance and then comparing the value of the calculated χ^2 to the actual χ^2 value from the tables. If the calculated value is greater, on the basis of the available data, the present model is inadequate and, consequently, appropriate design changes must be made.

In Section 7.3.3, three Portmanteau test statistics are defined for carrying out whiteness tests with nonseasonal ARIMA models. The Portmanteau statistic defined in [12.3.8] for use with SARIMA models is based upon the statistic given in [7.3.6] for employment with nonseasonal ARIMA models. The seasonal equivalent of the test statistic in [7.3.5] is

$$U_L = \eta'(\eta' + 2) \sum_{k=1}^{L} r_k^2(\hat{a})/(\eta' - k) \qquad [12.3.9]$$

This statistic is χ^2 distributed on $(L - P - p - Q - q)$ degrees of freedom.

Test for Periodic Correlation

When a SARIMA model is fitted to a seasonal hydrological series such as the average monthly flows for the Saugeen River plotted in Figure VI.1, it may not be able to describe the periodic or seasonal correlation that may be contained in the series. This is because the SARIMA model assumes that the correlation structure contained in the series is the same throughout the year. However, the correlation between riverflow values for July and August, for instance, may be quite different from the correlation between April and March. This fact is confirmed in Section 14.4 where a periodic autoregressive model is fitted to the average monthly Saugeen riverflows.

To test whether or not periodic correlation is contained in the residuals of a fitted SARIMA model, one can employ a statistical test presented by McLeod (1993). The periodic autocorrelation at lag k for season m may be written as

$$r_k^{(m)}(\hat{a}_{r,m}) = \frac{\sum_{r=1}^{n} \hat{a}_{r,m}\hat{a}_{r,m-k}}{\sqrt{\sum_{r=1}^{n} \hat{a}_{r,m}^2 \sum_{r=1}^{n} a_{r,m+k}^2}} \qquad [12.3.10]$$

where $\hat{a}_{r,m}$ is the estimated innovation or residual for the rth year and mth season, n is the number of years of seasonal data, and s is the number of seasons per year. Over the s seasons, the residual autocorrelations at lag one given by $r_1^{(m)}(\hat{a}_{r,m})$, $m = 1,2,\ldots,s$, are approximately jointly normally distributed with mean zero, diagonal covariance matrix, and $var(r_1^{(m)}(\hat{a}_{r,m})) = n^{-1}$. A diagnostic check for detecting periodic autocorrelation in the residuals of a fitted SARIMA model is given by

$$S = n \sum_{m=1}^{s} (r_1^{(m)}(\hat{a}_{r,m}))^2 \qquad [12.3.11]$$

which should be approximately χ^2 distributed on s degrees of freedom, if the model is adequate. When the calculated value for S is larger than that found in the tables for a given significance level, the calibrated model does not capture the periodic correlation.

Normality Tests

As pointed out in Section 7.4, many standard tests are available to check whether data are normally distributed. Additionally, the graph of the cumulative distribution of the residuals should appear as a straight line when plotted on normality paper if the residuals are normally distributed (Section 7.4.3). For instance, the residuals should not be significantly skewed or possess a significantly large kurtosis coefficient under the assumption that the residuals are normally distributed (Section 7.4.2).

Homoscedasticity Checks

Heteroscedasticity or changes in variance can arise in a number of different ways including:

1. the variance changes over time,
2. the magnitude of the variance is a function of the current level of the series.

Sensitive significance tests for checking for the presence of the above kinds of heteroscedasticity are described in detail in Section 7.5.

12.3.5 Summary

By following the identification, estimation and diagnostic checks stages of model construction, one can conveniently determine a reasonable SARMA or SARIMA model for describing a time series. These three construction stages are summarized in Figure 12.3.1. This approach follows the general model building procedure shown in Figure 6.3.2. The applications in the next stage demonstrate how easy it is to carry out this SARMA model building approach in practice.

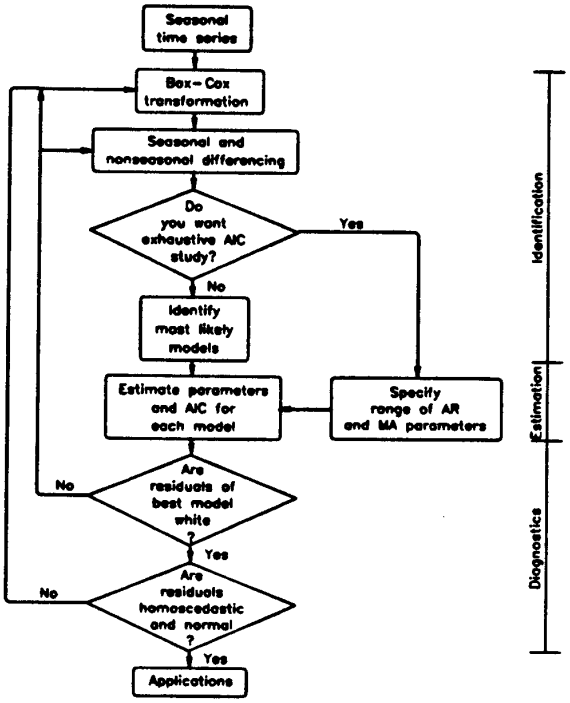

Figure 12.3.1 Constructing a SARIMA model.

12.4 APPLICATIONS

12.4.1 Introduction

To demonstrate how the model construction procedures of Section 12.3 are conveniently utilized in practice, SARIMA models are fitted to three seasonal time series. In the first application, an appropriate SARIMA model is designed for describing the average monthly water consumption series of Figure VI.2 while in the second case study a SARIMA model is fitted to the average monthly concentrations of atmospheric CO_2 displayed in Figure VI.3.

As pointed out in the Foreword to Part VI, the seasonal series displayed in Figures VI.2 and VI.3 constitute data sets for which the level of the series within each series increases with time. This nonstationary characteristic is clearly depicted in Figures VI.2 and VI.3 by the increasing trend around which the seasonal data fluctuate in sinusoidal patterns. However, for the average monthly flows of the Saugeen River at Walkerton, Ontario, shown in Figure VI.1, the mean and variance within a particular season across the years are more or less stationary and, consequently, there is no upward trend. In the third application, the best SARIMA model to fit to this series is determined. However, as mentioned in the Foreword to Section VI, the most appropriate types of models to fit to the average monthly Saugeen Riverflows are the deseasonalized and periodic models of Chapters 13 and 14, respectively. Although a calibrated SARIMA model for the Saugeen flows could be used for forecasting, it cannot be employed for simulation. This is because the SARIMA model is not designed for preserving stationarity within each season.

12.4.2 Average Monthly Water Useage

The average monthly water consumption series in millions of litres per day is available from the Public Utilities Commission (1989) of London from 1966 to 1988. When fitting a SARIMA model to this series, the first step is to obtain appropriate exploratory data analysis and identification graphs referred to in Section 12.3.2 and then to compare the information displayed on these graphs in order to design the SARIMA model.

As explained in Section 22.3, one of the most informative exploratory data analysis tools is simply a plot of the given series against time. A plot of the monthly water consumption versus time in Figure VI.2 certainly reveals important characteristics about the observations. As also noted in the foreword to part VI as well as the introduction to this section, the sinusoidal curve in Figure VI.2 indicates that the data are seasonal. Indeed, the demand for water is highest during the warmer summer months and lowest during the winter time. Moreover, the increasing linear trend component indicates that the data in each month of the year are increasing with time. In some instances a physical understanding of the phenomenon being analyzed allows for the incorporation of deterministic components into the model to account for seasonality and/or trends. For instance, seasonality may be modelled by a Fourier series, while trend might be accounted for by a polynomial. However, for the water demand data, a purely stochastic SARIMA model is fit to the data. Following the explanation of Section 4.6, the nonseasonal and seasonal differencing take care of the stochastic trend while the AR and MA parameters can describe the remaining dependence among the observations. Consequently, the SARIMA model stochastically accounts for the inherent properties of the data.

From the graph of the water demand series in Figure VI.2, it is not obvious that a data transformation is required. However, the variance appears to be increasing in the last two years of the water demand series. When a Box-Cox transformation is estimated for the SARIMA model identified below, the best transformation is found to be $\lambda = -0.75$.

The upward trend in Figure VI.2 points out that seasonal and perhaps also nonseasonal differencing are needed for removing the nonstationary behaviour. The nonstationarity of the raw data is also confirmed by the fact that the sample ACF plotted in Figure 12.4.1 dies off very slowly. The seasonality of the water demand data is reflected in the attenuating sine wave pattern in this figure.

Differencing the data once seasonally removes the nonstationarity contained in the original series. A graph of the seasonally differenced series in Figure 12.4.2 shows that differencing has eliminated the linear trend component as well as the sine wave. Also notice that the seasonally differenced series in Figure 12.4.2 has 12 fewer data points than the graph in Figure VI.2 because of the monthly differencing. In order to compare conveniently the original water consumption series in Figure VI.2 to the seasonally differenced version of the series in Figure 12.4.2, the two series can be plotted on a single graph. Figure 12.4.3 displays the bivariate trace plot for these graphs which shows Figure VI.2 as the lower graph and Figure 12.4.2 as the upper plot. To avoid clutter and permit easier interpretation of the results, the ordinate axis is omitted. One can clearly see from Figure 12.4.3 how seasonal differencing has removed trend and sinusoidal components.

The sample ACF, PACF, IACF and IPACF are displayed in Figures 12.4.4 to 12.4.7 for the seasonally differenced water demand series. Because none of these graphs attenuate slowly, the seasonally differenced data of Figures 12.4.2 and 12.4.3 (upper plot) are stationary. When

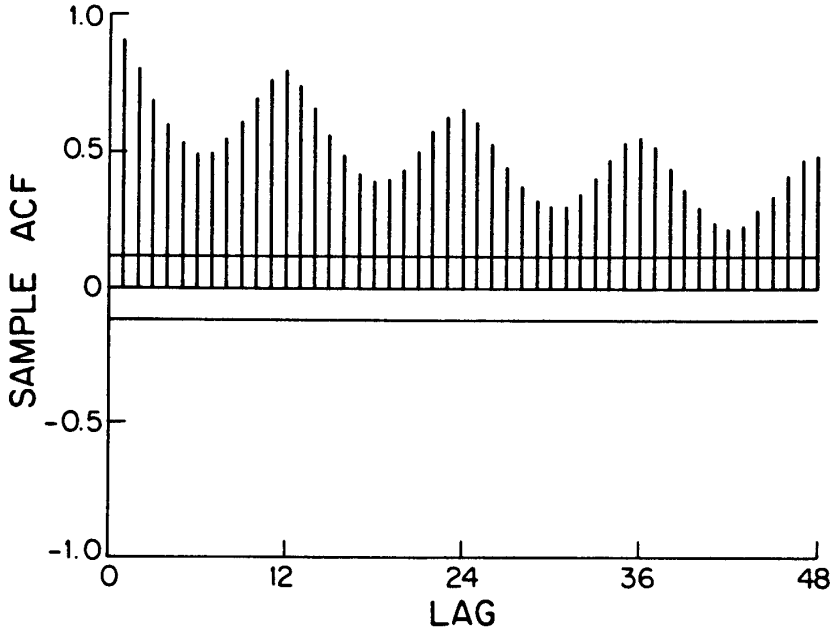

Figure 12.4.1. Sample ACF and 95% confidence limits for the average monthly water useage series from January, 1966, to December, 1988, for London, Ontario, Canada.

calculating the sample ACF in Figure 12.4.4, or, in general, when computing the ACF of any series that has been differenced, the mean of the w_t series in [12.2.8] is not removed. This procedure precludes missing any deterministic component that may still be present even after differencing.

In order to identify the number of AR and MA terms required in the model of the seasonally differenced monthly water useage data, the graphs of the sample ACF, the PACF, the IACF, and the IPACF that are shown in Figures 12.4.4 to 12.4.7, respectively, are interpreted simultaneously keeping in mind the main identification rules summarized in Table 12.3.1. Notice that both the sample ACF and IPACF have a significantly large value at lag 12. Moreover, because the sample PACF and IACF possess values that are decreasing in absolute magnitude at lags 12, 24, 36, and 48 (i.e. lags that are positive integer multiples of 12), this indicates the need for a seasonal MA term in the model.

Overall, the four identification graphs in Figures 12.4.4 to 12.4.7 complement one another in clearly pointing out the need for a seasonal MA parameter to include in the SARIMA model to fit to the seasonally differenced monthly water demand series. These graphs are also utilized for ascertaining which nonseasonal AR and MA parameters are needed. Because both the sample ACF and IPACF appear to die off across the first three or four lags, a nonseasonal AR parameter is required. Although the pattern is not strong, one could also interpret the sample PACF as attenuating during the first few lags. This indicates that a nonseasonal MA parameter may be needed in the model. Finally, notice that the sample IACF seems to cut off after the first

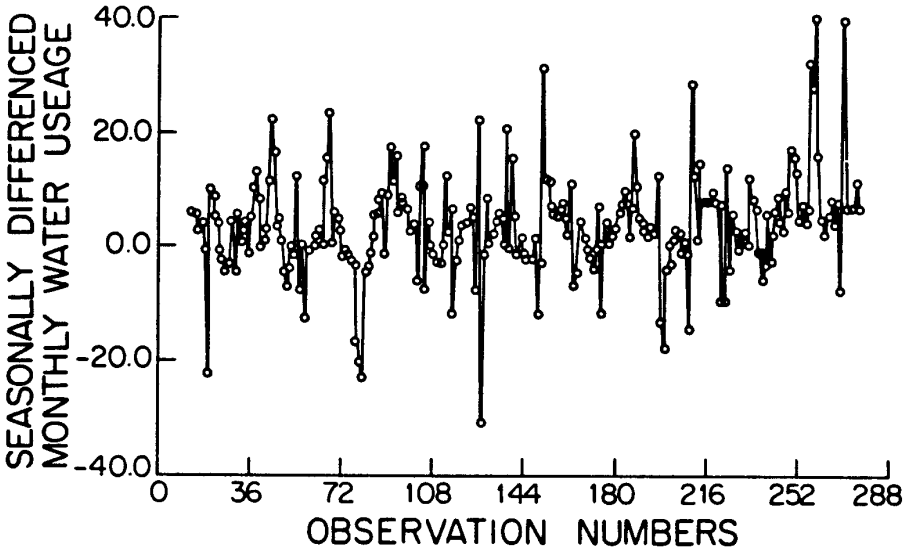

Figure 12.4.2. Graph of the seasonally differenced average monthly water useage series for London, Ontario, Canada.

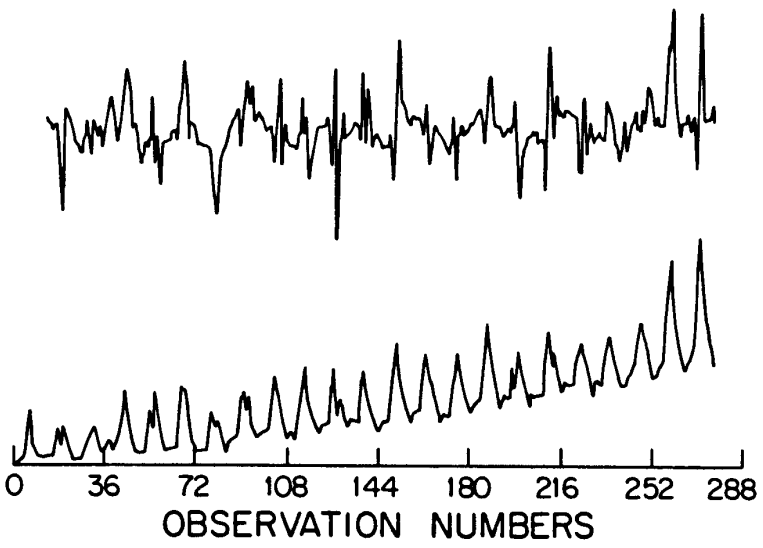

Figure 12.4.3. Bivariate trace plot of the average monthly water useage series (lower graph) for London, Ontario, Canada, from January, 1966, to December, 1988, and also the seasonally differenced water useage series (upper graph).

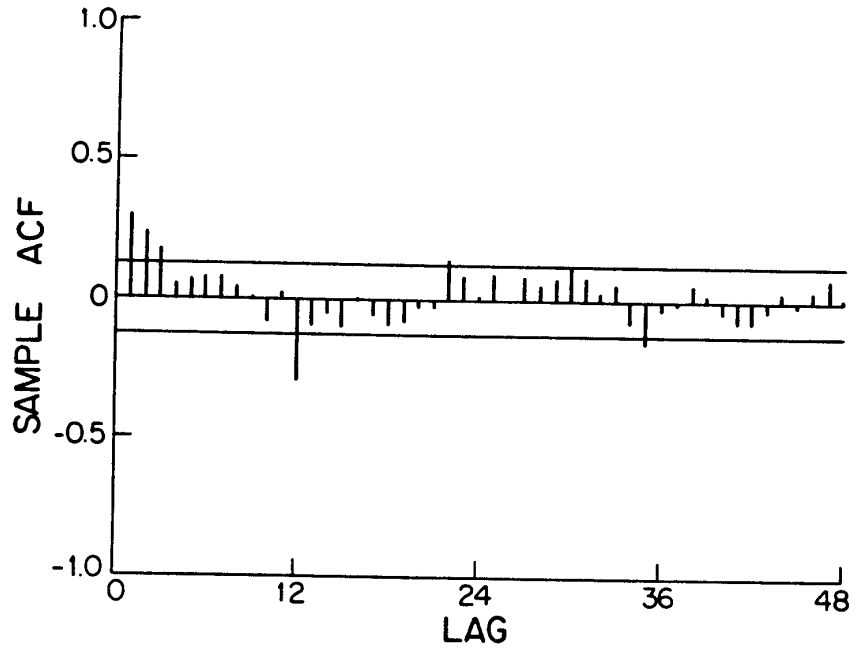

Figure 12.4.4. Sample ACF and 95% confidence limits for the seasonally differenced average monthly water useage series for London, Ontario, Canada.

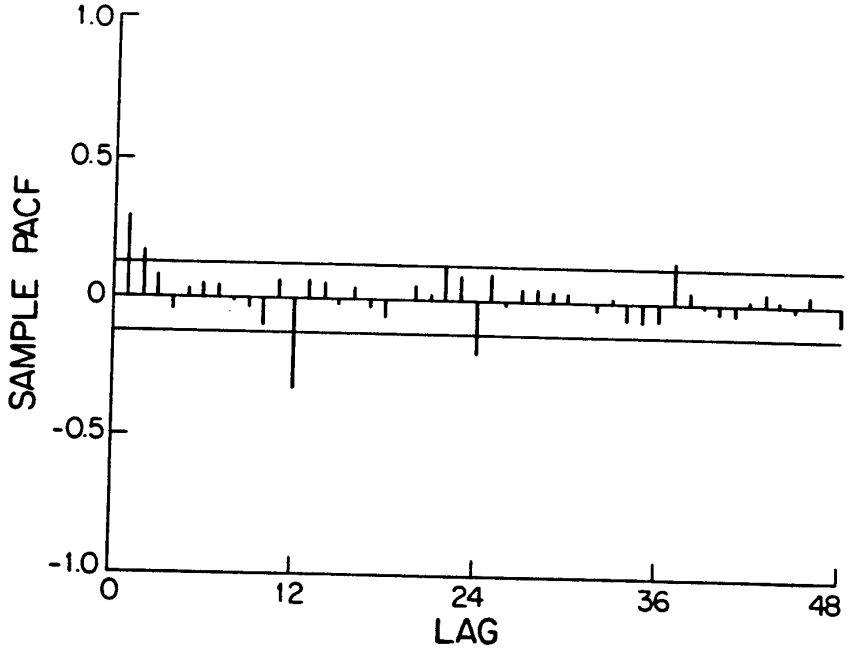

Figure 12.4.5. Sample PACF and 95% confidence limits for the seasonally differenced average monthly water useage series for London, Ontario, Canada.

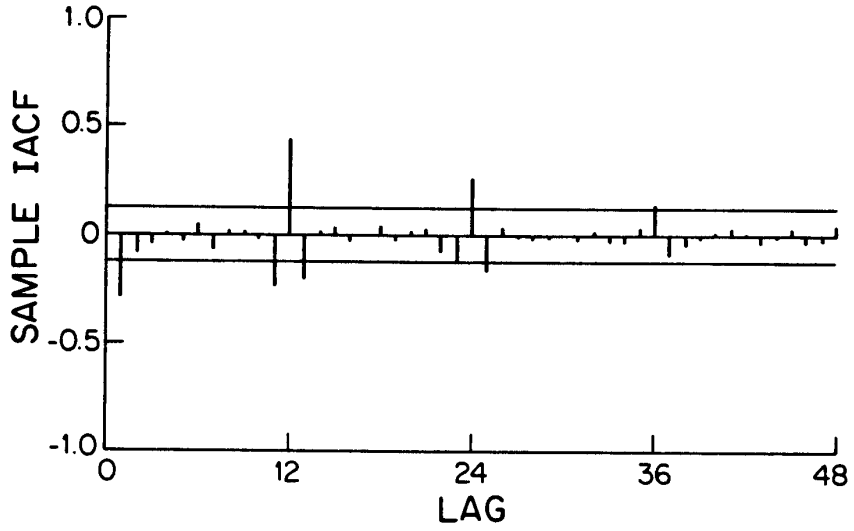

Figure 12.4.6. Sample IACF and 95% confidence limits for the seasonally differenced average monthly water useage series for London, Ontario, Canada.

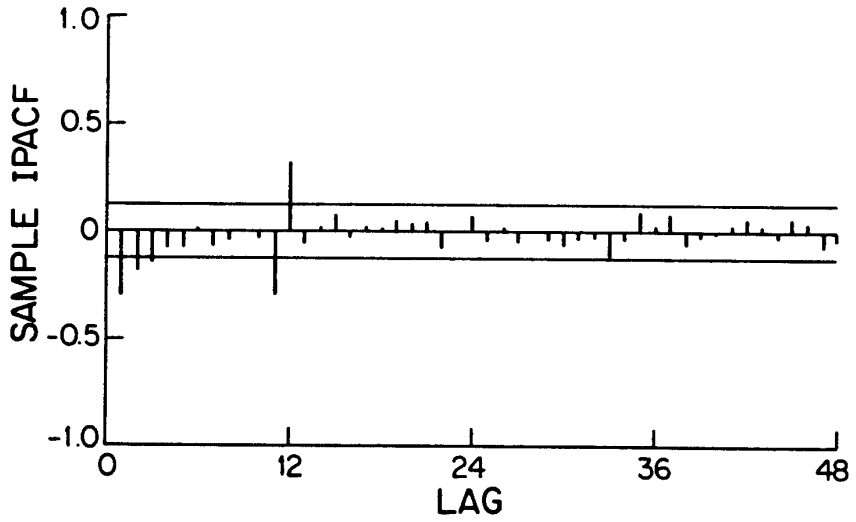

Figure 12.4.7. Sample IPACF and 95% confidence limits for the seasonally differenced average monthly water useage series for London, Ontario, Canada.

lag and, therefore, does not confirm the need for a nonseasonal MA parameter.

In summary, the identification plots given in Figures 12.4.4 to 12.4.7 indicate that an ARIMA(1,0,1)×(0,1,1)$_{12}$ or an ARIMA(1,0,0)×(0,1,1)$_{12}$ model should be fit to the water demand series. As mentioned earlier, a Box-Cox power transformation with $\lambda = -0.75$ is also required. When SARIMA models are calibrated for the seasonally differenced transformed water demand series, the SARIMA model possessing the lowest AIC calculated using [12.3.3] is the SARIMA(1,0,1)×(0,1,1)$_{12}$ model. Table 12.4.1 lists the parameter estimates and the SE's for this model while [12.4.1] gives the corresponding difference equation.

$$(1 - 0.618B)(1 - B^{12})z_t^{(\lambda)} = (1 - 0.297B)(1 - 0.851B^{12})a_t \qquad [12.4.1]$$

where the Box-Cox power parameter $\lambda = -0.75$ for water demand series, z_t.

Table 12.4.1. Parameter estimates and SE's for the SARIMA(1,0,1)×(0,1,1)$_{12}$ model with $\lambda = -0.75$ fitted to the water consumption series for London, Ontario, Canada.

Parameters	MLE's	SE's
ϕ_1	0.618	0.123
θ_1	0.297	0.150
Θ_1	0.851	0.032
σ_a^2	1.86×10^{-6}	

Notice in Table 12.4.1 that the estimate for the nonseasonal MA parameter is about twice its SE and is, therefore, just barely significantly different from zero. On the other hand, the parameter estimates for ϕ_1 and Θ_1 are many times larger than their corresponding SE's. Recall that the need for incorporating both ϕ_1 and Θ_1 into the model are clearly indicated by the identification graphs.

The calibrated SARIMA model in [12.4.1] passes diagnostic checks for whiteness, normality and homoscedasticity refereed to in Section 12.3.4. Figure 12.4.8, for example, of the RACF for the fitted SARIMA model shows that the model residuals are white. Except for lag 22, all of the RACF values fall within the 95% confidence interval. This large value at lag 22 is probably due to chance alone and could not be removed by including other model parameters. One would expect that there is one chance in twenty that a value could fall outside the 95% confidence limits. Additionally, this large value does not occur at crucial lags such as 1, 12, 24, and 36. Other diagnostic checks reveal that both the homoscedasticity and the normality assumption for the residuals are fulfilled. Therefore, on the basis of the information used, the chosen SARIMA model [12.4.1] adequately models the monthly water useage data.

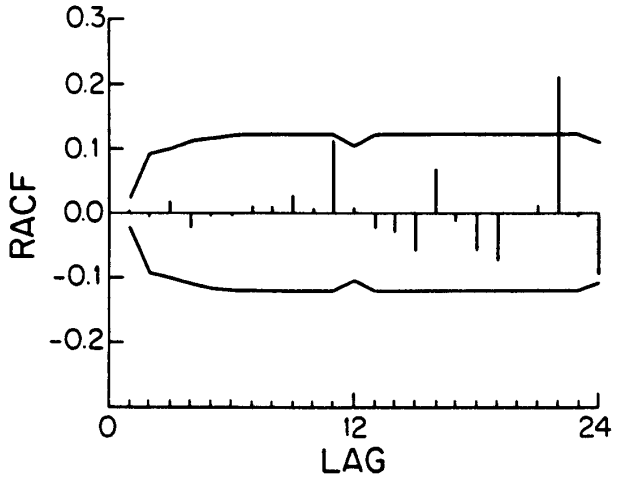

Figure 12.4.8. RACF and 95% confidence limits for the SARIMA$(1,0,1)\times(0,1,1)_{12}$ model with $\lambda = -0.75$ fitted to the average monthly water useage series for London, Ontario, Canada.

12.4.3 Average Monthly Atmospheric Carbon Dioxide

Bacastow and Keeling (1981) as well as Keeling et al. (1982) list average monthly concentrations of atmospheric CO_2 measured at the Mauna Loa Observatory on the Island of Hawaii. Figure VI.3 displays a plot of these observations from January, 1965, to December, 1980. Because carbon dioxide is a principal greenhouse gas that could cause global warming, the monitoring and analysis of CO_2 data such as the series in Figure VI.3 is of wide interest to environmental scientists, political decision makers and, indeed, the general public. As pointed out by Bacastow and Keeling (1981), the increase per year in carbon dioxide concentration in the atmosphere is one of the important observations for understanding the carbon cycle.

Blue (1991) reports upon the work of Wahlen et al. (1991) who are taking measurements of CO_2 in the air of bubbles in the GISP 2 (Greenland Ice Sheet Project 2) ice core using a dry extraction technique and tunable diode laser absorption spectroscopy. The CO_2 record spans the years from 1530 to 1940 and includes parts of the little ice age, a time of abnormally cold temperatures in Europe during the *16th* and *18th* centuries. Wahlen et al. (1991) have found that there were no significant changes in CO_2 concentrations during the little ice age. However, since about 1810, the CO_2 concentrations have started to increase due to industrialization and other related land use changes such as deforestation and urbanization. Moreover, the date of the onset of increasing CO_2 in about 1810 in the GISP 2 ice core is similar to that discovered in the Siple core from Antarctica (Neftel et al., 1982) and also in another Antarctic ice core analyzed by Pearman et al. (1986). Finally, the data retrieved from the Greenland ice core GISP 2 are

consistent with observations from Mauna Loa.

Figure VI.3 clearly depicts an increasing linear trend in the monthly CO_2 levels over the years. The sinusoidal curve wrapped around the trend demonstrates that the data are sinusoidal with larger values occurring in May or June of each year. Because of the nonstationarity in Figure VI.3, differencing is required to remove trends and hence create a stationary series. Figure 12.4.9 shows a bivariate trace plot for which the original CO_2 series of Figure VI.3 is given in the lower half of the graph and the stationary series is plotted at the top when the given series is differenced both seasonally (i.e. $D = 1$) and nonseasonally ($d = 1$). Notice from the top series the way in which differencing has removed the increasing trend as well as the distinct seasonality shown in the original lower series. Also, because of the differencing, the upper series is 13 ($d + D = 1 + 12 = 13$) data points shorter than the original lower series.

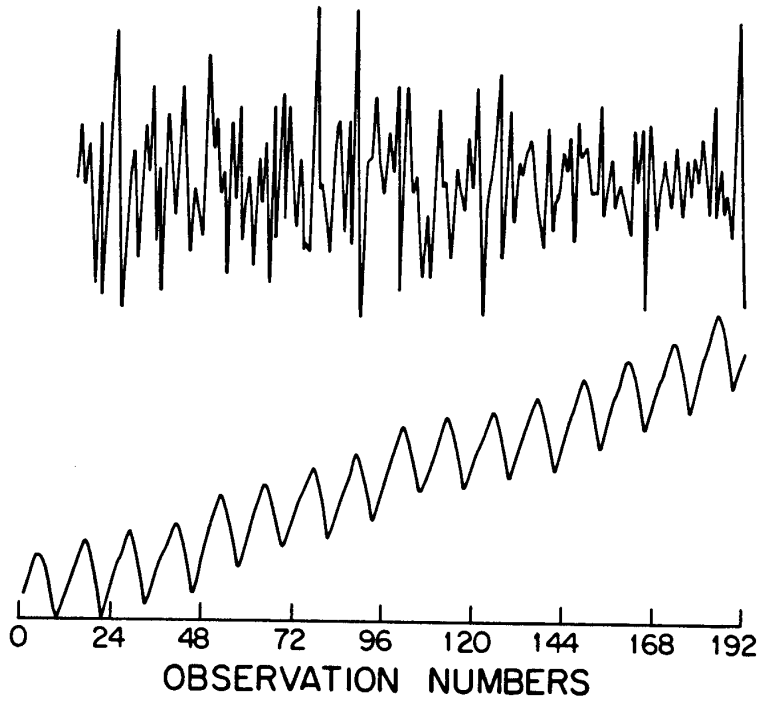

Figure 12.4.9. Bivariate trace plot of the average monthly concentrations of atmospheric CO_2 (mole fraction in ppm) (lower graph) measured at Mauna Loa Observatory in Hawaii from January, 1965, to December, 1980, and also the nonseasonally and seasonally differenced CO_2 series (upper graph).

To discover the AR and MA parameters required to model the top series in Figure 12.4.9, one can simultaneously examine the sample ACF, PACF, IACF and IPACF shown in Figures 12.4.10 to 12.4.13, respectively. The significantly large values of the sample ACF and IPACF at lag 12 means that a seasonal MA parameter is needed in the model. This finding is also confirmed by the fact that the sample PACF and IACF attenuate at lags that are positive integer multiples of 12 (i.e. lags 12, 24, 36 and 48). The slightly large values at lag one for the sample ACF points out that a nonseasonal MA parameter may be needed. The fact that the sample IACF possesses attenuating values after lags 1, 12, 24 and 36 may indicate that a nonseasonal MA parameter is required. Consequently, the most appropriate model to fit to the monthly CO_2 data is probably a SARIMA$(0,1,1)\times(0,1,1)_{12}$ model. When comparing other possible SARIMA models such as the SARIMA$(1,1,1)\times(0,1,1)_{12}$ and $(1,1,0)\times(0,1,1)_{12}$ to the aforementioned model, the AIC value calculated using [12.3.3] is lowest for the SARIMA$(0,1,1)\times(0,1,1)_{12}$ model.

Table 12.4.2 lists the MLE's and SE's for the parameters of the SARIMA$(0,1,1)\times(0,1,1)_{12}$ model fitted to the monthly CO_2 data set. In difference equation form, this calibrated SARIMA model is written as

$$(1-B)(1-B^{12})z_t = (1-0.336B)(1-0.831B^{12})a_t \qquad [12.4.2]$$

where z_t represents the average monthly CO_2 series value at time t.

The estimated model in [12.4.2] provides a reasonable fit to the CO_2 series according to diagnostic checks for whiteness, normality and constant variance described in Sections 7.3 to 7.5, respectively. Because all of the values for the RACF fall within the 95% confidence limits in Figure 12.4.14, the model residuals are not significantly correlated and, hence, are white. Moreover, the calibrated model performs reasonably well with respect to normality and homoscedasticity checks. For example, the value of the test statistic for changes in variance depending on the current level (see Section 7.5.2) is -2.232 with a SE of 1.296. Because the test statistic falls within two SE's of zero, one can argue that the statistic is not significantly different from zero and, hence, the residuals are homoscedastic.

12.4.4 Average Monthly Saugeen Riverflows

The average monthly flows for the Saugeen River at Walkerton, Ontario, Canada, are displayed in Figure VI.1. When modelling monthly riverflow series it is usually necessary to take natural logarithms of the data to alleviate problems with heteroscedasticity and/or non-normality in the model residuals. Therefore, the Box-Cox parameter λ is set equal to zero in [12.2.1] for the Saugeen flows. Following this, an examination of the sample ACF for the logarithmic data reveals that the data has to be differenced once seasonally in order to remove seasonal nonstationarity. By studying the properties of the sample ACF, PACF, IACF and IPACF graphs for the seasonally differenced logarithmic Saugeen time series, it is found that the best design is a SARIMA$(1,0,1)\times(0,1,1)_{12}$ model. Diagnostic checks for the residuals of the calibrated model indicate that the model provides an adequate fit to the data. However, it is found that the periodic autocorrelation test statistic in [12.3.11] for the fitted SARIMA has a significantly large value of 59.6 on twelve degrees of freedom. Consequently, the SARIMA model residuals possess significant periodic correlation. Therefore, in Section 14.4 a periodic autoregressive model is fitted to the logarithmic average monthly Saugeen riverflows in order to

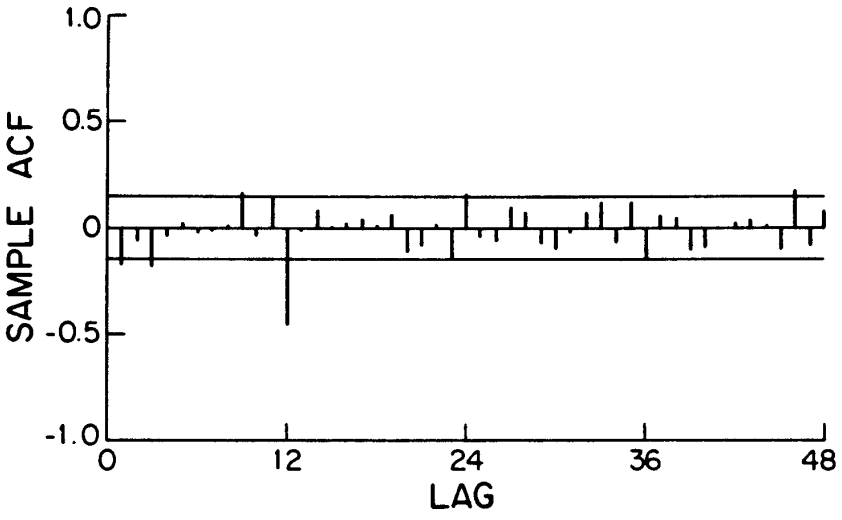

Figure 12.4.10. Sample ACF and 95% confidence limits for the differenced ($d = D = 1$) average monthly atmospheric CO_2 concentrations (mole fraction in ppm) measured at Mauna Loa Observatory in Hawaii.

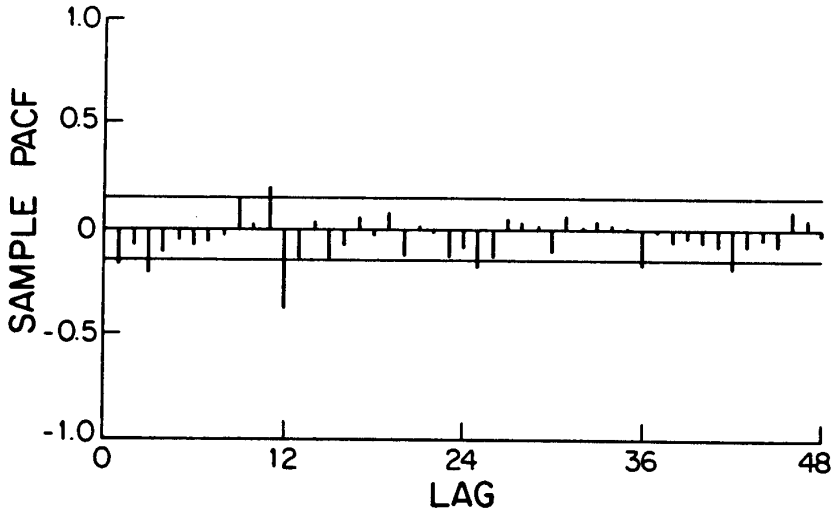

Figure 12.4.11. Sample PACF and 95% confidence limits for the differenced ($d = D = 1$) average monthly atmospheric CO_2 concentrations (mole fraction in ppm) measured at Mauna Loa Observatory in Hawaii.

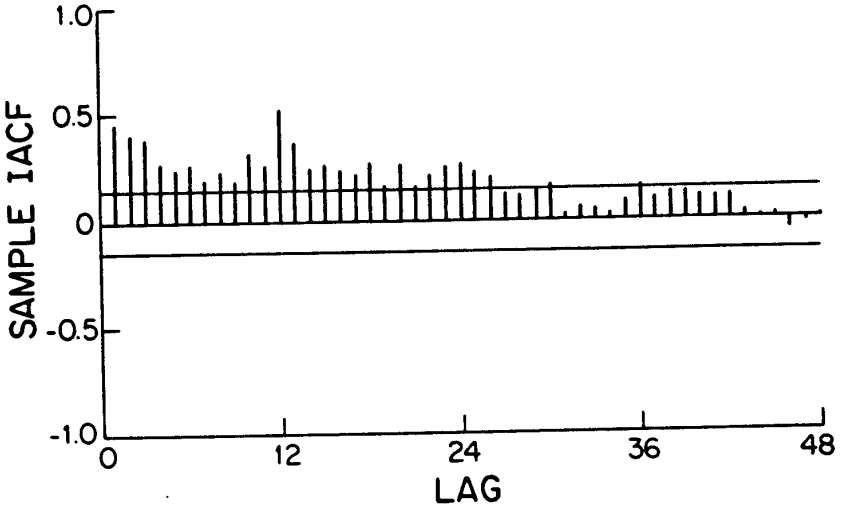

Figure 12.4.12. Sample IACF and 95% confidence limits for the differenced ($d = D = 1$) average monthly atmospheric CO_2 concentrations (mole fraction in ppm) measured at Mauna Loa Observatory in Hawaii.

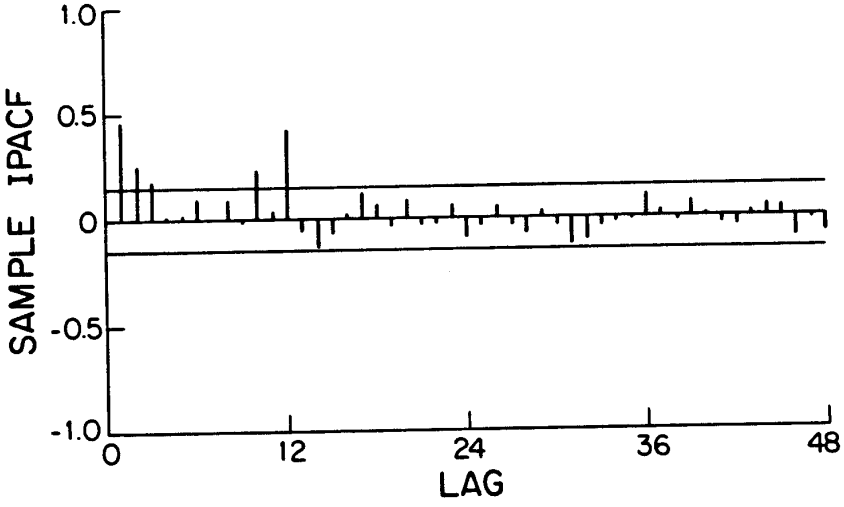

Figure 12.4.13. Sample IPACF and 95% confidence limits for the differenced ($d = D = 1$) average monthly atmospheric CO_2 concentrations (mole fraction in ppm) measured at Mauna Loa Observatory in Hawaii.

Table 12.4.2. Parameter estimates and SE's for the SARIMA$(0,1,1)\times(0,1,1)_{12}$ model fitted to the average monthly atmospheric CO_2 concentrations measured at Mauna Loa Observatory in Hawaii.

Parameters	MLE's	SE's
θ_1	0.336	0.070
Θ_1	0.831	0.041
σ_a^2	1.01×10^{-1}	

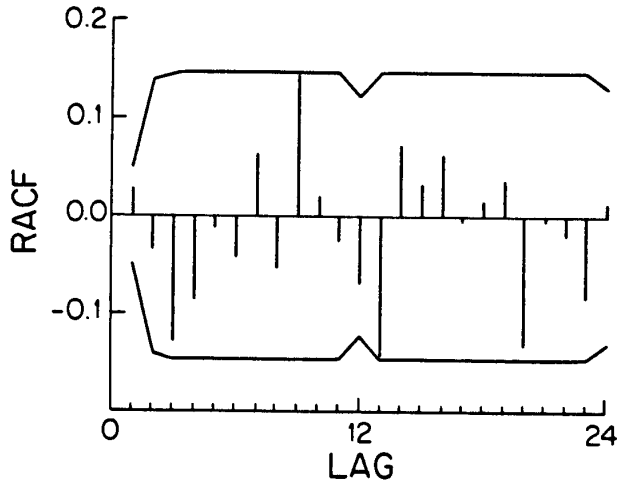

Figure 12.4.14. RACF and 95% confidence limits for the SARIMA$(0,1,1)\times(0,1,1)_{12}$ model fitted to the average monthly atmospheric CO_2 concentrations (mole fraction in ppm) measured at Mauna Loa Observatory in Hawaii.

properly model the periodic correlation.

Equation [12.3.3] can be employed for calculating the AIC of the SARIMA model fitted to the Saugeen flows. The value of the AIC is found to be 3435.43. As shown in Chapters 13 and 14 for the values of the AIC determined for the deseasonalized and periodic models, respectively, the estimate of the AIC is much higher for the calibrated SARIMA model. This is because deseasonalized and periodic models are specifically designed for preserving certain kinds of stationarity within each season of the average monthly Saugeen riverflows.

12.5 FORECASTING AND SIMULATING WITH SARIMA MODELS

After fitting the SARIMA model in [12.2.7] to a given seasonal time series, the calibrated model can be employed for forecasting and simulation. In Chapter 8, minimum mean square error forecasts are defined and procedures are presented for calculating MMSE forecasts for non-seasonal ARMA (Chapter 3) and ARIMA (Chapter 4) models. Section 15.2.2 explains how MMSE forecasts are conveniently determined for a SARIMA model when the model is written using the generalized form in [12.2.12]. If the data have been transformed using a Box-Cox transformation, one can employ the procedure explained in Section 15.2.2 as well as Section 8.2.7 to determine the forecasts in the original untransformed domain.

In Section 12.4.2 of the previous section, the most appropriate SARIMA model to fit to the average monthly water useage series displayed in Figure VI.2 is found to be the SARIMA$(1,0,1)\times((0,1,1)_{12}$ model with $\lambda = -0.75$ written in [12.4.1]. Figure 12.5.1 depicts the MMSE forecasts for the water demand series from January 1989 until December 1990 that are calculated for the fitted SARIMA model by following the procedure given in Section 15.2.2. The seasonal characteristics of these forecasts can be clearly seen as a sinusoidal pattern on the right hand side of Figure 12.5.1 that fall within their 90% probability limits.

The model for the SARIMA$(0,1,1)\times(0,1,1)_{12}$ model fitted to the average monthly concentrations of atmospheric CO_2 is written in [12.4.2]. The MMSE forecasts for this model are displayed on the right hand side of Figure 12.5.2 along with the 90% probability limits. As is also the case in Figure 12.5.1, the forecasts follow the seasonal sinusoidal shape exhibited by the historical observations.

When simulating with a SARIMA model, the first step is to write the model in unfactored form as in [12.2.11]. Next, WASIM1 or WASIM2 explained in Sections 9.4 and 9.5, respectively, is employed to determine simulated data for the w_t series in [12.2.11]. Subsequently, the algorithm of Section 9.5.2 is utilized for integrating the systematic w_t values to obtain the simulated $z_t^{(\lambda)}$ series. Finally, the inverse Box-Cox transformation in [9.6.2] is invoked to procure the corresponding z_t synthetic data in the untransformed domain. Parameter uncertainty can be entertained by employing the WASIM3 algorithm of Section 9.7.

12.6 CONCLUSIONS

The SARIMA and SARMA models defined in [12.2.7] and [12.2.9], respectively, are designed for modelling time series which exhibit nonstationarity both within and across seasons. An inherent advantage of the SARIMA family of models is that relatively few model parameters are required for describing these types of time series. As demonstrated by the three applications in Section 12.4, the model construction techniques of Section 12.3 can be conveniently and expeditiously implemented in practice for designing, calibrating and checking SARIMA models. Many other applications of SARIMA models to water resources and environmental time series can be found in the literature. For instance, Irvine and Eberhardt (1992) fit SARIMA models to lake level time series.

As explained in Section 12.5 and also Section 15.2.2, a calibrated SARIMA model can be employed for forecasting and simulation. In Chapter 15, the results of forecasting experiments demonstrate that the deseasonalized and periodic models of Chapters 13 and 14, respectively, forecast better than SARIMA models when forecasting monthly riverflow time series. This is

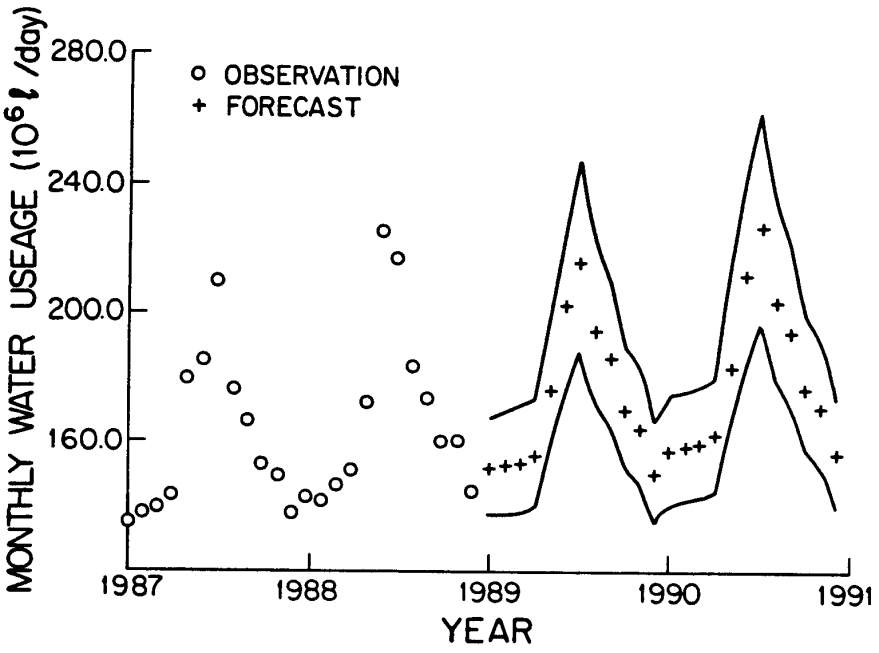

Figure 12.5.1. MMSE forecasts along with their 90% probability intervals for the SARIMA$(1,0,1)\times(0,1,1)_{12}$ model with $\lambda = -0.75$ fitted to the average monthly water useage (million litres per day) series from 1966 to 1988 for London, Ontario, Canada.

because the SARIMA model is designed for modeling the type of series shown in Figures VI.2 and VI.3 rather than the one in Figure VI.1.

If there were more than one seasonal cycle in a series, other sets of seasonal operators could be incorporated into the SARIMA model to handle this situation. Furthermore, one could also easily design a seasonal FARMA (fractional ARMA model) by directly extending the definitions in Chapter 11 for nonseasonal FARMA models to the seasonal case.

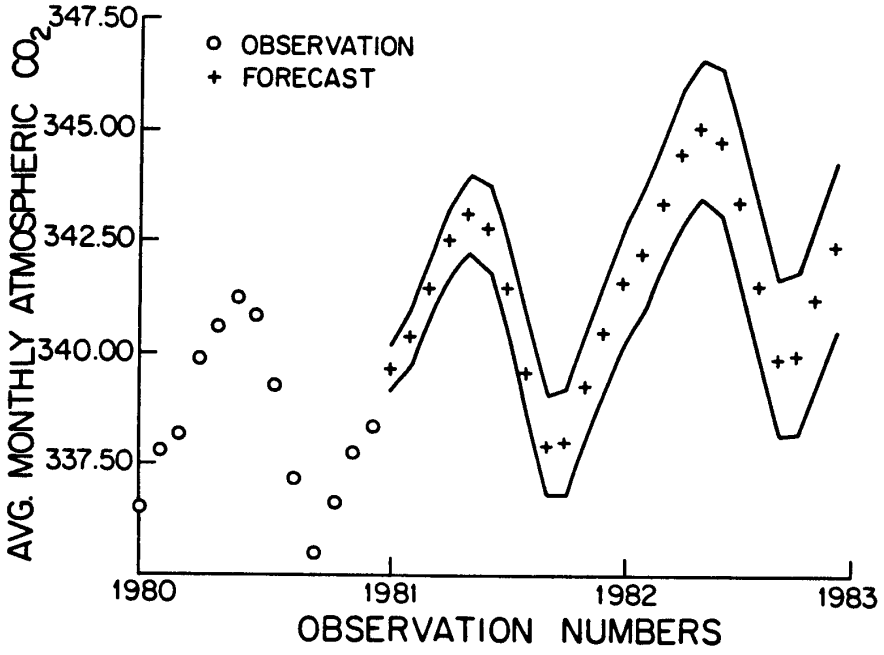

Figure 12.5.2. MMSE forecasts along with their 90% probability intervals for the SARIMA$(0,1,1)\times(0,1,1)_{12}$ model fitted to the average monthly water concentrations of atmospheric CO_2 (mole fraction in ppm) measured at Mauna Loa observatory in Hawaii.

APPENDIX A12.1

DESIGNING

MULTIPLICATIVE SARIMA MODELS

USING THE ACF

The multiplicative SARMA$(p,q)\times(P,Q)_s$ model for fitting to the stationary w_t series in [12.2.8] is defined in [12.2.9]. The theoretical ACF of w_t is given by

$$\rho_k = \gamma_k/\gamma_0, \quad k = 0,1,2,\cdots$$

where $\gamma_k = E[w_t w_{t-k}]$. Using a result of Godolphin (1977, [3.4]), it follows that for s large enough

$$\rho_{i\pm js} \simeq \rho_i(u_t)\rho_j(U_t) \quad \text{for } i = 0,1,\ldots,s/2 \text{ and } j = 0,1,\cdots, \qquad \text{[A12.1.1]}$$

where $\rho_{i\pm js}$ is the ACF of w_t at lag $i\pm js$ and $\rho_i(u_t)$ and $\rho_j(U_t)$ are the values of the ACF at lags i and j of u_t and U_t in the models $\phi(B)u_t = \theta(B)a_t$ and $\Phi(B^s)U_t = \Theta(B^s)a_t$, respectively. Godolphin (1977, [3.4]) demonstrates using a vector representation that [A12.1.1] holds exactly when $p = P = 0$ and $s > 2q$. As shown at the end of this appendix, this result may also be derived using the autocovariance generating function (Box and Jenkins, 1976, p. 81). In any case, the general approximation follows from the fact that a SARMA$(p,q)\times(P,Q)_s$ model can be approximated by a $(0,q)\times(P,Q)_s$ model for suitable Q when s is large enough.

A graphical interpretation, along similar lines to that suggested by Hamilton and Watts (1978) for the PACF may be given. First, the regular nonseasonal component pattern is defined as the ACF of u_t plus repetitions centered about the seasonal lags $0,s,2s,\cdots$. Figure A12.1.1 shows the regular component pattern corresponding to a $(1,0)(P,Q)_{12}$ model with $\phi_1 = 0.6$. Next, the seasonal component pattern is determined from the ACF of U_t as illustrated in Figure A12.1.2 for the case of a $(p,q)(1,0)_{12}$ model with $\Phi_1 = 0.6$. Then, the approximation to the ACF of w_t is the product of the regular and seasonal component patterns. Figure A12.1.3 shows the resulting approximation for the $(1,0)(1,0)_{12}$ model with $\phi_1 = \Phi_1 = 0.6$.

In conclusion, the general interpretation presented in this appendix is useful for seasonal model identification and does not appear to have been pointed out previously. For some applications, the employment of this procedure can simplify the identification process.

Generating Function Proof of Godolphin's Result

Theorem: If $s > 2q$, the autocovariance function, γ_k, of a $(0,q)\times(P,Q)_s$ model may be expressed for nonnegative lags as,

$$\gamma_{js\pm i} = \begin{cases} \gamma_i(u_t)\gamma_j(U_t) & \text{for } 0 \le i \le q \text{ and } 0 \le j \le Q, \\ 0, & \text{otherwise,} \end{cases}$$

where $\gamma_i(u_t)$ and $\gamma_j(U_t)$ denote the autocovariance functions of the processes $\phi(B)u_t = \theta(B)a_t$ and $\Phi(B)U_t = \Theta(B)a_t$.

Proof: For convenience, it may be assumed that $var(a_t) = 1$.

$$\Gamma(B) = \theta(B)\theta(B^{-1})/[\Phi(B)\Phi(B^{-1})]$$

$$= \sum_{k=-\infty}^{\infty} \Gamma_k B^k$$

Then the autocovariance generating function of the $(0,q)(P,Q)_s$ model may be written as

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k$$

$$= \theta(B)\theta(B^{-1})\Gamma(B^s)$$

$$= \left[\left(\sum_{l=0}^{q} \sum_{i=0}^{q-l} \theta_i \theta_{i+l} \right) (B^l + B^{-l}) \right] \Gamma(B^s),$$

where $\theta_0 = -1$. Thus, provided $s > 2q$, the coefficient of B^{js+i} $(i = 0, \cdots, q)$ is

$$\gamma_{js+i} = \Gamma_j \sum_{l=0}^{q-i} \theta_l \theta_{l+i}$$

Similarly, the coefficients of B^{js+i} and B^{js-i} can be shown to be equal.

APPENDIX A12.2

MAXIMUM LIKELIHOOD ESTIMATION FOR SARMA MODELS

McLeod and Salas (1983) provide an algorithm for calculating an approximation to the likelihood function of the multiplicative SARMA model in [12.2.9]. Their algorithm specifically takes advantage of the multiplicative structure of the nonseasonal and seasonal AR and MA operators in the SARMA model. The conditional, unconditional or iterated unconditional method of Box and Jenkins (1976) may be used in the algorithm of McLeod and Salas (1983) in conjunction with an approximation to the determinant term (see McLeod (1977) and also Appendix 6.1) to obtain an accurate and highly efficient algorithm. In fact, McLeod and Salas (1983) point out that other competing algorithms become computationally infeasible when the seasonal period s becomes much larger than 12, as in the cases of half-monthly ($s = 24$), weekly ($s = 52$) or daily ($s = 365$) time series.

In this appendix, the theory behind the algorithm of McLeod and Salas (1983) is outlined. The reader can refer to the paper of McLeod and Salas (1983) for a more detailed description of the theory and method of application as well as a listing of the Fortran computer program for the algorithm.

The SARMA $(p,q)\times(P,Q)_s$ model for fitting to a series w_t of length $\eta' = \eta - d - sD$ is defined in [12.2.9]. Let the vector of model parameters be given by

$$\beta = (\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_p, \Phi_1, \Phi_2, \ldots, \Phi_p, \Theta_1, \Theta_2, \ldots, \Theta_Q)$$

Although the SARMA model may be considered as a special case of the ARMA (p^*,q^*) model in [12.2.14] by taking $p^* = sP$, $q^* = q + sQ$, $\phi^*(B) = \Phi(B^s)\phi(B)$ and $\theta^*(B) = \Theta(B^s)\theta(B)$, a more efficient estimation algorithm can be developed utilizing the multiplicative structure of the SARMA model.

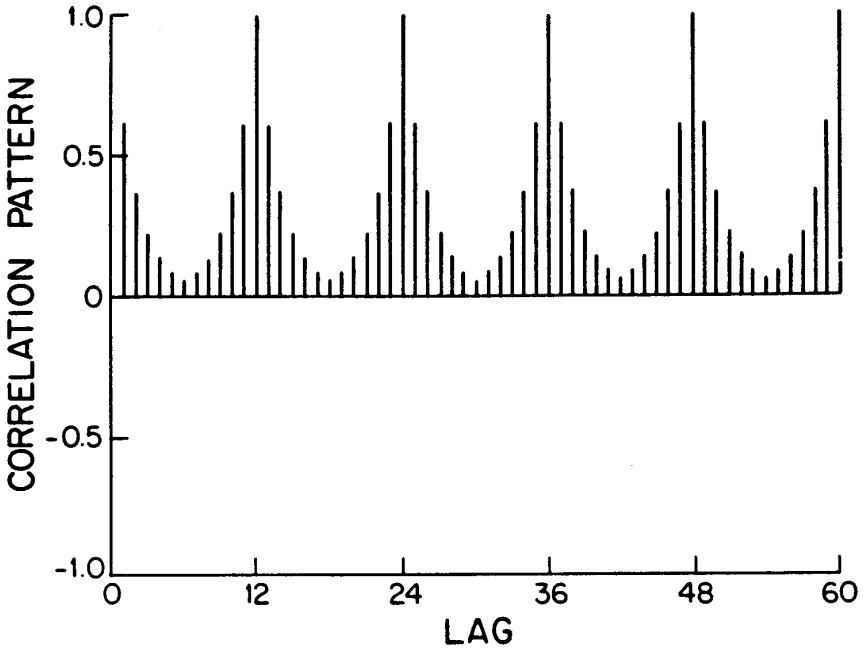

Figure A12.1.1. A regular nonseasonal component pattern of a SARIMA(1,0)$(P,Q)_{12}$ model with $\phi_1 = 0.6$.

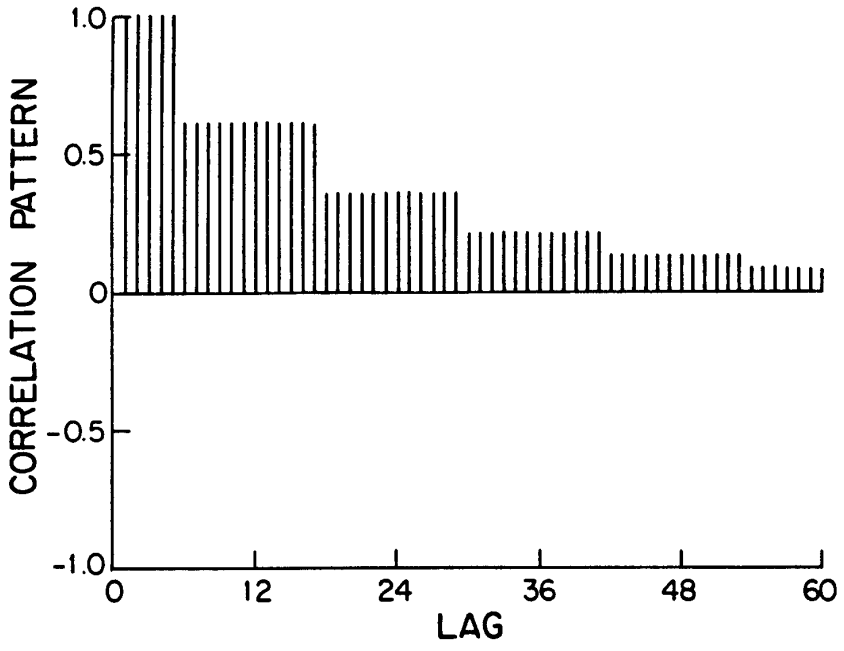

Figure A12.1.2. Seasonal component pattern of a SARMA$(p,q)\times(1,0)_{12}$ with $\Phi_1 = 0.6$.

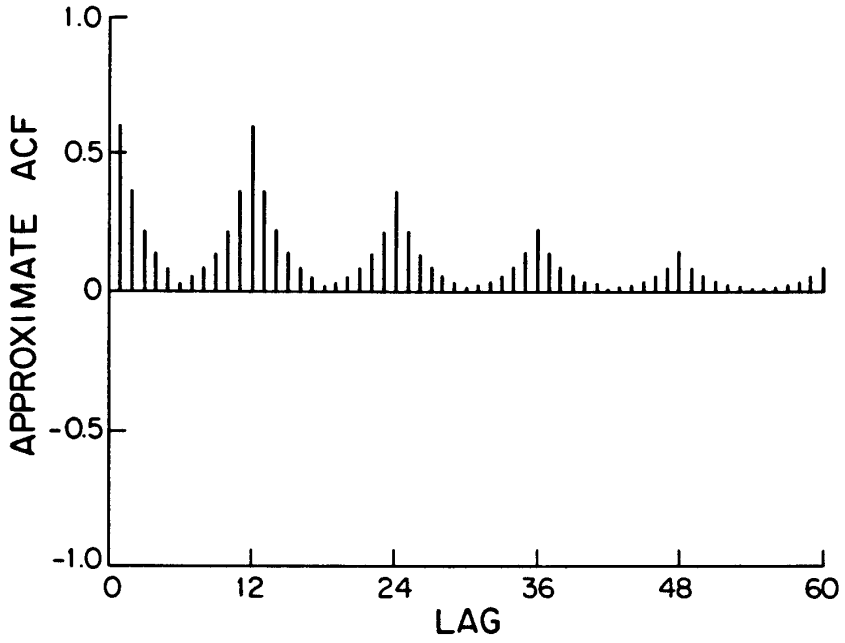

Figure A12.1.3. Approximate ACF of a SARMA(1,0)(1,0)$_{12}$ model with $\phi_1 = 0.6$ and $\Phi_1 = 0.6$.

Given the observations w_t, t=1,2, $\cdots$,η', the exact log-likelihood function maximized over the variance, σ_a^2, of the innovation, a_t, may be written, apart from an arbitrary constant,

$$\log L(\beta) = -\eta' \log(S_m/\eta')/2 \qquad \text{[A12.2.1]}$$

where S_m, the modified sum of squares is

$$S_m = S[M_{\eta'}(p,q,P,Q,s)]^{-1/\eta'}. \qquad \text{[A12.2.2]}$$

S represents the unconditional sum of squares of Box and Jenkins (1976) defined by

$$S = \sum_{t=-\infty}^{n} [a_t]^2, \qquad \text{[A12.2.3]}$$

where [.] denotes expectation given $w_1, w_2, \cdots, w_{\eta'}$.

The evaluation of S by the iterative unconditional sum of squares method may involve two types of truncation error. First, the infinite sum in [A12.2.3] is replaced by

$$S = \sum_{t=1-T}^{n} [a_t]^2 \qquad \text{[A12.2.4]}$$

for suitably large T. Theoretically, T should be chosen so that

$$\gamma_0/\sigma_a^2 - \sum_{i=0}^{T} \psi_i^2 < e_{t01} \qquad \text{[A12.2.5]}$$

where $\gamma_0 = var(w_t)$, ψ_i is the coefficient of a_{t-i} in the infinite MA representation of [12.2.9] and e_{t01} is an error tolerance. Thus, if the model contains a AR factor with roots near the unit circle, a fairly large T might be necessary. In practice,

$$T = q + sQ + 20(p + sQ) \qquad \text{[A12.2.6]}$$

is often sufficient. The other truncation error involves terminating the iterative procedure used to calculate $[a_t]$. Several iterations may be required to obtain convergence when the model contains a MA factor with roots near the unit circle. However, sufficient accuracy is usually obtained on the first evaluation.

McLeod (1977) suggests that the term $M_n(p,q,P,Q,s)$ be replaced by $m(p,q,P,Q,s)$, given by

$$m(p,q,P,Q,s) = M(p,q)[M(P,Q)]^s, \qquad \text{[A12.2.7]}$$

where $M(p,q)$ is defined for any ARMA(p,q) model as

$$M(p,q) = M_p^2 M_q^2 / M_{p+q} \qquad \text{[A12.2.8]}$$

where the terms M_p, M_q and M_{p+q} are defined in terms of the auxiliary autoregressions, $\phi(B)v_t = a_t$ and $\theta(B)u_t = a_t$ and the left-adjoint autoregression $\phi(B)\theta(B)y_t = a_t$. For the autoregression, $\phi(B)v_t = a_t$, M_p is the determinant of the $p \times p$ matrix with (i,j) entry

$$\sum_{k=1}^{\min(i,j)} \phi_{i-k}\phi_{j-k} - \phi_{p+k-i}\phi_{p+k-j} \qquad \text{[A12.2.9]}$$

and similarly for the other autoregressions. The $p \times p$ matrix defined by [A12.2.9] is called the Schur matrix of $\phi(B)$. Pagano (1973) has shown that a necessary and sufficient condition for stationarity of an autoregression is that its Schur matrix be positive-definite (see Section 3.2.2). Thus, calculation of $m(p,q,P,Q,s)$ also provides a check on the stationarity and invertibility conditions and so during estimation the parameters may be constrained to the admissible region. Modified Cholesky decomposition is used to evaluate $M(p,q)$.

To obtain MLE's for the model parameters, the modified sum of squares must be minimized by using a standard optimization algorithm (see Section 6.2.3). McLeod and Salas (1983) describe in detail how the backforecasting method of Box and Jenkins (1976, Ch. 7) for ARMA models can be efficiently adapted for employment with SARMA models by making use of their multiplicative structure.

PROBLEMS

12.1 Select an average monthly riverflow time series that has at least ten years of data. Employ a suitable set of time series programs such as the McLeod-Hipel Package referred to in Section 1.7 to help you perform the following tasks.

(a) Examine a graph of the observations plotted over time and comment upon the overall statistical characteristics of the data. If necessary use other exploratory data analysis tools (see Sections 1.2.4, 5.3.2, 12.3.2 and 22.3) to study the statistical properties of your series.

(b) Based upon the discussion given at the start of Part VI and elsewhere in the chapter, which type of seasonal model do you feel is most appropriate to fit to the time series?

(c) Utilizing the techniques of Section 12.3, follow the three stages of model construction to develop the most appropriate SARIMA model to fit to this series. Show and explain all of your modelling results at each model building stage.

12.2 Carry out question 12.1 for a seasonal socio-economic time series of your choice. For example, you may wish to study a quarterly water or electrical demand series.

12.3 Execute question 12.1 for a seasonal water quality time series that has no missing values.

12.4 Select a seasonal meteorological time series for answering the questions given in question 12.1.

12.5 Write a SARIMA$(2,1,3)\times(1,1,2)_4$ model that is fitted to a series z_t with $\lambda = 0.5$ in the following forms:

(a) difference equation format given in [12.2.7],

(b) unfactored form in [12.2.10],

(c) generalized style as in [12.2.12],

(d) random shock form given in [12.2.15],

(e) and inverted format written in [12.2.17].

12.6 Derive the theoretical ACF for the SARMA$(p,q)\times(P,Q)$ model in [12.2.9].

12.7 For a SARIMA$(1,1,2)\times(1,1,1)_{12}$ model fitted to a series with $\lambda = 0.5$, calculate the MMSE forecasts for $l = 1,2,\ldots,24$.

12.8 For the same SARIMA model given in 12.7, describe the steps for simulating 10,000 sequences of length 120 for this model. Where necessary, use equations to explain how calculations are carried out.

12.9 Select one of the calibrated SARIMA models that you fitted to a seasonal time series in problems 12.1 to 12.4. Calculate and plot the MMSE forecasts as well as the 90% probability intervals using this model for lead times from 1 to 25 where s is the seasonal length. Comment upon your results.

12.10 Choose one of the calibrated SARIMA models that you fitted to a seasonal data set in one of the first four questions. Simulate and plot three synthetic sequences of length $8s$ for the SARIMA model where s is the seasonal length. Compare these simulated sequences to the original series and discuss your findings.

12.11 Outline the procedure of Box et al. (1987) for estimating the trend in a seasonal time series that can be described using a SARIMA model. Comment upon the advantages and drawbacks of their approach.

12.12 Briefly describe how the tests of Li (1991) work for determining the orders of differencing required for modelling a seasonal time series. Comment upon the usefulness of the differencing tests.

REFERENCES

CARBON DIOXIDE DATA

Bacastow, R. B. and Keeling, C. D. (1981). Atmospheric carbon dioxide concentration and the observed airborne fraction. In Bolin, B., editor, *Carbon Cycle Modelling*, pages 103-112. John Wiley, Chichester.

Blue, C. (1991). CO_2 in glacial ice gives clues to historical atmospheres. *EOS, Transactions of the American Geophysical Union*, 72(36):379.

Keeling, C. D., Bacastow, R. B. and Whorf, T. P. (1982). Measurements of the concentration of carbon dioxide at Mauna Loa Observatory, Hawaii. In Clark, W. C., Editor, *Carbon Dioxide Review 1982*, pages 377-385. Clarendon Press, Oxford.

Neftel, A., Oeschger, H., Schwander, J., Stauffer, B. and Zumbrunn, R. (1982). Ice core sample measurements give atmospheric CO_2 content during the past 40,000 years. *Nature*, 295:216-219.

Pearman, G. I., Etheridge, D., de Silva, F. and Fraser, P. J. (1986). Evidence of changing concentrations of atmospheric CO_2, N_2O, and CH_4 from bubbles in Antarctic ice. *Nature*, 320:248-250.

Wahlen, M., Allen, D., Deck, B. and Herchenroder, A. (1991). Initial measurements of CO_2 concentrations (1530 to 1940 AD) in air occluded in the GISP2 ice core from Central Greenland. *Geophysical Research Letters*, 18(8):1457-1460.

RIVERFLOW AND WATER CONSUMPTION DATA

Environment Canada (1977). Historical streamflow summary, Ontario. Technical report, Inland Waters Directorate, Water Resources Branch, Ottawa, Canada.

Public Utilities Commission (1989). Water useage data for the city of London, Ontario. Technical report, Public Utilities Commission, P.O. Box 2700, London, Ontario.

SARIMA MODELLING

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19:716-723.

Bartlett, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society*, Series B, 8:27-41.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, Series B, 26:211-252.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day, Oakland, California, revised edition.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of the residual autocorrelations in autoregressive integrated moving average models. *Journal of the American Statistical Association*, 65:1509-1526.

Box, G. E. P., Pierce, D. A. and Newbold, P. (1987). Estimating trend and growth rates in seasonal time series, *Journal of the American Statistical Association*, 82(397):276-282.

Cleveland, W. S. (1972). The inverse autocorrelations of a time series and their applications. *Technometrics*, 14(2):277-298.

Godolphin, E. J. (1977). On the autocorrelation structure for seasonal moving average models and its implications for the Cramer-Wold decomposition. *Journal of Applied Probability*, 14:785-794.

Granger, C. W. J. and Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society*, Series B, 38(2):189-203.

Hamilton, D. C. and Watts, D. G. (1978). Interpreting partial autocorrelation functions of seasonal time series models. *Biometrika*, 65(1):135-140.

Hipel, K. W., McLeod, A. I. and Lennox, W. C. (1977). Advances in Box-Jenkins modelling, 1, Model construction. *Water Resources Research*, 13(3):567-579.

Irvine, K. N. and Eberhardt, A. J. (1992). Multiplicative seasonal ARIMA models for Lake Erie and Lake Ontario water levels, *Water Resources Bulletin*, 28(2):385-396.

Li, W. K. (1991). Some Lagrange multiplier tests for seasonal differencing, *Biometrika*, 78(2):381-387.

Li, W. K. and McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *Journal of the Royal Statistical Society*, Series B, 43(2):231-239.

McLeod, A. I. (1977). Improved Box-Jenkins estimators. *Biometrika*, 64(3):531-534.

McLeod, A. I. (1978). On the distribution of residual autocorrelations in Box-Jenkins models. *Journal of the Royal Statistical Society*, Series B, 40(3):296-302.

McLeod, A. I. (1984). Duality and other properties of multiplicative seasonal autoregressive-moving average models. *Biometrika*, 71:207-211.

McLeod, A. I. (1993). Parsimony, model adequacy and periodic correlation in time series forecasting. *The International Statistical Review*, 61(3).

McLeod, A. I., Hipel, K. W. and Lennox, W. C. (1977). Advances in Box-Jenkins modelling, 2, Applications. *Water Resources Research*, 13(3):577-586.

McLeod, A. I. and Salas, P. R. H. (1983). An algorithm for approximate likelihood calculation of ARMA and seasonal ARMA models. *Journal of the Royal Statistical Society*, Series C (Applied Statistics), 32:211-223.

Ozaki, T. (1977). On the order determination of ARIMA models. *Journal of the Royal Statistical Society*, Series C (Applied Statistics), 26(3):290-301.

Pagano, M. (1973). When is an autoregressive scheme stationary? *Communications in Statistics*, 1(6):533-544.