PART IV

FORECASTING

AND

SIMULATION

Within Part II of the book, two useful classes of nonseasonal models are defined and some of their theoretical properties are derived. In particular, the **ARMA** family of models of Chapter 3 are defined for fitting to stationary time series while the **ARIMA** class of models presented in Chapter 4 are designed for use with nonstationary data sequences. A sensible and systematic approach for fitting these and other kinds of models to a given data set is described in Part III. More specifically, by following the identification, estimation and diagnostic check stages of **model construction** explained in Chapters 5 to 7, respectively, one can develop the most appropriate model to describe the data set being studied.

A particular ARMA or ARIMA model which has been fitted to a time series can serve a variety of **useful purposes**. For example, the calibrated model provides an economic means of **encoding** the basic statistical properties of the time series into a few model parameters. In the process of carrying out the model building procedure, one obtains a **better understanding** about the key statistical characteristics of the data set. Besides the **insights** which are always gained when fitting a model to a time series, there are two important types of applications of time series models which are in widespread use by practitioners. These application areas are forecasting and simulation. The objectives of Part IV are to explain how ARMA and ARIMA models can be used for forecasting and simulation, and furnish case studies for demonstrating how forecasting and simulation are executed in practice.

The general purpose of **forecasting** or **prediction** is to provide the best estimates of what will happen at specified points in time in the future. Based upon the model fitted to a series and the most recent observations, one can obtain what are called **minimum mean square error forecasts** of future observations. Because forecasting is concerned with using the fitted model to extrapolate the time series into the future, it is often called **extrapolation**. Moreover, since forecasting, prediction or extrapolation provides an estimate of the future behaviour of a system, it is essential in the **operation** and **control** of the system. For example, forecasts for a riverflow series could be used for deciding upon the long range operating rules of a large reservoir. Forecasting can also be employed for **model discrimination**. When models from a variety of different classes are fitted to time series, one can select the model which provides the most accurate forecasts. The theory and practice of forecasting with nonseasonal models are presented in Chapter 8.

The overall objective of **simulation** is to use a fitted model to generate possible future values of a time series. These simulated or **synthetic sequences** can be used in two main ways. Firstly, simulated sequences can be utilized in **engineering design**. For instance, when designing a reservoir complex for generating hydroelectrical power, one can use both the historical flows and simulated data for obtaining the most economical design. Simulated sequences are employed in the design process because when the reservoir comes into operation the future flows will never be exactly the same as the historical flows. Therefore, one wishes to subject tentative designs to a wide variety of stochastically possible flow scenarios. Secondly, simulation can be employed for studying the **theoretical properties** of a given model. In many cases, it is very difficult or, for practical purposes, impossible, to determine precise analytical results for a given theoretical property of the model. When this is the situation, simulation can be used for obtaining the theoretical results to a specified desired level of accuracy. The theory and practice of simulating with nonseasonal models are explained in Chapter 9.

The forecasting and simulation techniques presented in the next two chapters are explained in terms of ARMA and ARIMA models. However, these methods can be easily extended for use with other models such as the different seasonal models of Part VI and the transfer function-noise models of Part VII. Table 1.6.3 lists the locations in the book where contributions to forecasting and simulation are given for a wide range of time series models.

CHAPTER 8

FORECASTING WITH NONSEASONAL MODELS

8.1 INTRODUCTION

In the design, planning and operation of water resources systems, one often needs good estimates of the future behaviour of key hydrological variables. For example, when operating a reservoir to serve multiple purposes such as hydroelectrical power generation, recreational uses and dilution of pollution downstream, one may require forecasts of the projected flows for upcoming time periods. The objective of forecasting is to provide accurate predictions of what will happen in the future.

In practical applications, forecasts are calculated after the most appropriate time series model is fitted to a given sequence of observations. Figure III.1 summarizes how a model is developed for describing the time series by following the identification, estimation and diagnostic check stages of model construction. Figure 6.3.1 outlines how an automatic selection criterion such as the AIC (Akaike information criterion) can be utilized in model building. After obtaining a calibrated model, one can calculate forecasts for one or more time steps into the future. Figure 8.1.1 displays the overall procedure for obtaining forecasts. Notice that the original data set may be first transformed using an appropriate transformation such as the Box-Cox transformation in [3.4.30]. Whatever the case, subsequent to constructing a time series model to fit to the series by following the procedures of Part III, one can use the calibrated model and the most recent observations to produce forecasts in the transformed domain. If, for example, an original data set of annual riverflows were first transformed using natural logarithms, then the forecasts from the ARMA model fitted to the logarithmic data would be predictions of the logarithmic flows. As indicated in Figure 8.1.1, one would have to take some type of inverse data transformation of the forecasted flows in the transformed domain in order to obtain forecasts in the original domain. These forecasts could then be used for an application such as optimizing the operating rules of a reservoir.

When calculating forecasts, one would like to obtain the most accurate forecasts possible. However, the question arises as to how one quantifies this idea of accuracy. One useful criterion for defining accuracy is to use what is called minimum mean square error. The theoretical definition of what is meant by *minimum mean square error forecasts* and the method of calculating them for ARMA and ARIMA models are presented in the next section. In addition, the method for calculating confidence limits for the forecasts is described.

Forecasting can be used as an approach for *model discrimination*. A variety of time series models can be fitted to the first portion of one or more time series and then used to forecast the remaining observations. By comparing the accuracy of the forecasts from the models, one can determine which set of models forecasts the best. In Section 8.3, *forecasting experiments* are carried out for deciding upon the best types of models to use with yearly natural time series.

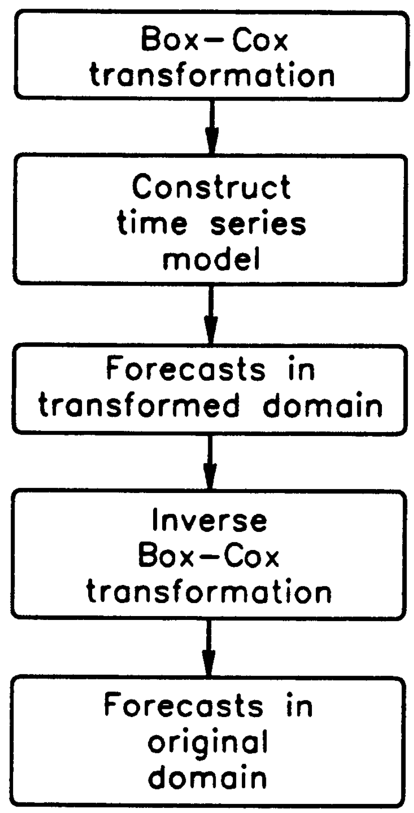

Figure 8.1.1. Overall procedure for obtaining forecasts.

When comparing one step ahead forecasts, some statistical tests are described for determining if one model forecasts significantly better than another.

Chapter 8 deals with forecasting using nonseasonal ARMA and ARIMA models. Forecasting with other kinds of models is described in other chapters of the book. In Chapter 15, the procedures for forecasting with three types of *seasonal models* are presented. Forecasting experiments are also given in Sections 15.3 and 15.4 for comparing the forecasting abilities of different seasonal models. Procedures for combining forecasts from distinctly different models in order to obtain overall better forecasts are described in Section 15.5. Similar approaches could also be used for combining forecasts from different nonseasonal models. Finally, Chapter 18 describes how one can obtain forecasts using a *transfer function-noise model*. As explained in Chapter 17, a transfer function-noise model is a time series model that can describe situations where there is a single output and multiple inputs. For example, the output series may be riverflows whereas the input or covariate series are precipitation and temperature measurements. Table 1.6.3

summarizes where material on forecasting can be found in the book.

Besides hydrology, forecasting experiments have been carried out in other disciplines to compare the forecasting ability of models. In economics, one important forecasting study was completed by Newbold and Granger (1974). In their investigation these authors used one hundred and six economic time series to compare three types of forecasting models. The time series were split into two parts and ARIMA, Holt-Winters, and stepwise autoregressive models were fitted to the first portion of the data. The three models were then used to forecast the remainder of the data for various lead times. The forecasting ability of the three models was judged on the basis of the mean squared error (MSE) of the forecasts. Newbold and Granger (1974) found that the ARIMA forecasting procedure clearly outperformed the other two methods for short lead times but the advantage decreased for increasing lead time.

Madridakis et al. (1982) reported on a recent forecasting competition. The forecasting ability of over twenty models was tested using 1001 time series. The time series were of different length, type (i.e., monthly, quarterly, and annual) and represented data ranging from small firms to nations. Different forecast horizons were considered and several criterion were employed to compare the forecasts from the various models. In general, no one specific model produces superior forecasts for all types of data considered. However, some improvement may be achieved if the forecaster selected certain classes of models for forecasting specific types of data.

Because of the great import of forecasting in water resources engineering as well as many other disciplines, there have been many research papers, conference proceedings and books written on forecasting. Most of the water resources and time series analysis books referred to in Chapter 1 of this book contain chapters on forecasting. The Hydrological Forecasting Symposium (International Association of Hydrological Sciences, 1980) held in Oxford, England, certainly confirms the usefulness of forecasting in hydrology. For forecasting in economics, readers may wish to refer to texts listed in the references under economic forecasting at the end of Chapter 1. Within this book, recent practical developments for forecasting in water resources engineering are presented.

8.2 MINIMUM MEAN SQUARE ERROR FORECASTS

8.2.1 Introduction

Let z_t represent a known value of a time series observed at time t. For convenience of explanation, assume for now that the data have not been transformed using an appropriate data transformation. In Section 8.2.7, it is explained how the Box-Cox transformation in [3.4.30] is taken into account when forecasting. As shown in Figure 8.2.1, suppose that the observations are known up until time t. Given an ARMA or ARIMA model that is fitted to the historical series up to time t, one wishes to use this model and the most recent observation to forecast the series at time $t+l$. Let the forecast for the unknown observation, z_{t+l}, be denoted by $\hat{z}_t(l)$, since one is at time t and would like to forecast l steps ahead. The time t is referred to as the *origin time* for the forecast while l is the *lead time* which could take on values of $l = 1,2,....$ Consequently, in Figure 8.2.1, the forecasts from origin t having lead times of $l = 1$, 2, and 3, are denoted by $\hat{z}_t(1)$, $\hat{z}_t(2)$, and $\hat{z}_t(3)$, respectively. The forecast, $\hat{z}_t(1)$, at lead time 1, is called the *one step ahead forecast* and is frequently used in forecasting experiments for discriminating among competing models.

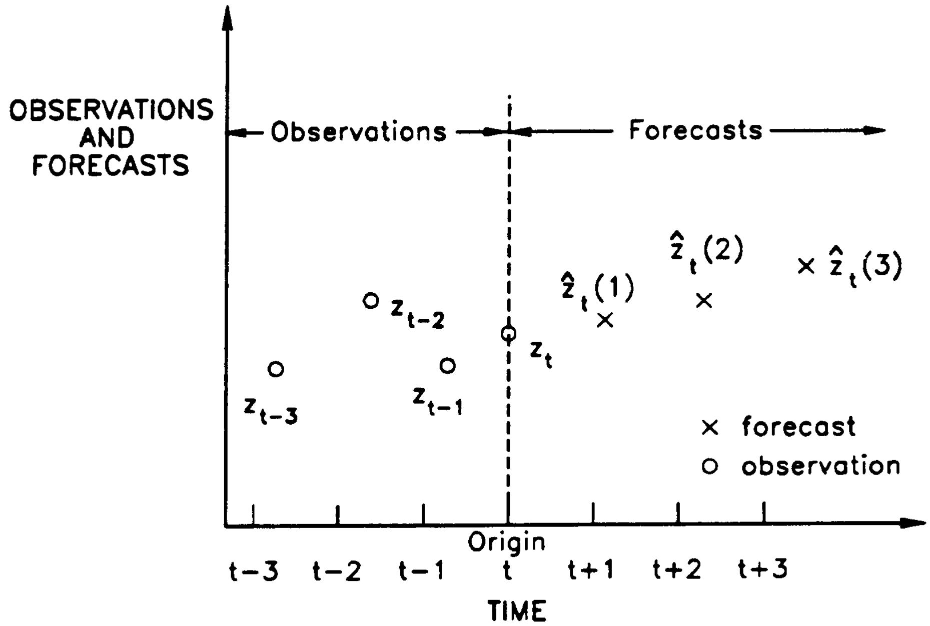

Figure 8.2.1. Forecasts from origin t.

One would like to produce forecasts which are as close as possible to what eventually takes place. Another way to state this is that one would like to minimize the forecast errors. This is because larger forecast errors can lead to poor decisions which in turn can cause more excessive costs than would be necessary. For example, if a hydroelectric complex were operated inefficiently because of poor forecasts, the utility could lose large sums of money.

To appreciate what is meant by *forecast errors*, refer once again to Figure 8.2.1. After the observation at time t+1 becomes known, the *one step ahead forecast error* from origin t is calculated as

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

Likewise, the forecast errors for lead time two and three are determined, respectively, as

$$e_t(2) = z_{t+2} - \hat{z}_t(2)$$

$$e_t(3) = z_{t+3} - \hat{z}_t(3)$$

In general, the forecast error at lead time l is given as

$$e_t(l) = z_{t+l} - \hat{z}_t(l) \quad l = 1,2,... \qquad [8.2.1]$$

Decision makers would like to minimize forecast errors in order to keep the costs of their decisions as low as possible. However, when calculating forecasts for lead times $l = 1,2,\ldots,k$, how should one define the forecast error that should be minimized? For example, one approach is to minimize the *mean error* given by

$$\bar{e} = \frac{1}{k}\sum_{l=1}^{k} e_t(l) \qquad [8.2.2]$$

Another would be to minimize the *mean absolute error (MAE)* written as

$$MAE = \frac{1}{k}\sum_{l=1}^{k} |e_t(l)| \qquad [8.2.3]$$

A third alternative is to minimize the *mean square error (MSE)* defined as

$$MSE = \frac{1}{k}\sum_{l=1}^{k} e_t(l)^2 \qquad [8.2.4]$$

One could easily define other criteria for defining forecast errors to be minimized. For instance, one could weight the forecast errors according to their time distance from origin t and then use these weighted errors in any of the above types of overall errors. As pointed out in the next subsection, minimum mean square error forecasts possess many attractive properties that have encouraged their widespread usage in practical applications.

8.2.2 Definition

As explained in Sections 3.4.3 and 4.3.4 for ARMA and ARIMA models, respectively, these two classes of models can be written in any of the three equivalent forms:

1. difference equation form as originally defined,
2. random shock format (i.e. as pure MA model)
3. inverted form (i.e. as a pure AR model).

Any one of these three forms of the model can be used for calculating the type of forecasts defined in this section. However, for presenting the definition of what is meant by a *minimum mean square error (MMSE) forecast* when using an ARMA or ARIMA model the random shock model, is most convenient to use.

From [3.4.18] or [4.3.9], at time t, the random shock model is written as

$$\begin{aligned} z_t &= \psi(B)a_t \\ &= (1 + \psi_1 B + \psi_2 B^2 + \cdots)a_t \\ &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \end{aligned} \qquad [8.2.5]$$

where $\psi(B) = (1 + \psi_1 B + \psi_2 B^2 + \cdots)$ is the random shock or infinite MA operator for which ψ_i is the ith parameter and a_t is the innovation sequence distributed as NID$(0,\sigma_a^2)$. When standing at time $t + l$, the random shock model is given as

$$\begin{aligned} z_{t+l} &= \psi(B)a_{t+l} \\ &= (1 + \psi_1 B + \psi_2 B^2 + \cdots)a_{t+l} \\ &= a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \cdots \end{aligned}$$

$$+\psi_l a_t + \psi_{l+1}a_{t-1} + \psi_{l+2}a_{t-2} + \cdots \qquad [8.2.6]$$

For simplifying the explanation, the mean of the series is omitted in the above two equations and ensuing discussions. However, when there is a nonzero mean for z_t, all of the upcoming results remain exactly the same.

Suppose that standing at origin t, one would like to make a forecast $\hat{z}_t(l)$ of z_{t+l} which is a linear function of current and previous observations $z_t, z_{t-1}, z_{t-2}, \ldots,$. This in turn implies that the forecast is a linear function of current and previous innovations $a_t, a_{t-1}, a_{t-2}, \ldots,$. Using all of the information up to time t and the random shock form of the model in [8.2.6], let the best forecast at lead time l be written as

$$\hat{z}_t(l) = \psi_l^* a_t + \psi_{l+1}^* a_{t-1} + \psi_{l+2}^* a_{t-2} + \cdots \qquad [8.2.7]$$

where the weights $\psi_l^*, \psi_{l+1}^*, \psi_{l+2}^*, \ldots,$ are to be determined. Notice that the innovations $a_{t+1}, a_{t+2}, \ldots, a_{t+l}$, and their corresponding coefficients are not included in [8.2.7] since they are unknown.

The theoretical definition of the mean square error of the forecast is defined as $E[z_{t+l} - \hat{z}_t(l)]^2$. By replacing z_{t+l} and $\hat{z}_t(l)$ by the expressions given in [8.2.6] and [8.2.7], respectively the mean square error is expanded as

$$E[z_{t+l} - \hat{z}_t(l)]^2 = E[(a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \cdots) - (\psi_l^* a_t + \psi_{l+1}^* a_{t-1} + \psi_{l+2}^* a_{t-2} + \cdots)]^2$$

After expanding the right hand side by squaring and then taking the expected value of each term, the equation is greatly simplified because of the fact that

$$E[a_t a_{t-j}] = \begin{cases} 0, & j \neq 0 \\ \sigma_a^2, & j = 0 \end{cases}$$

More specifically, the equation reduces to

$$E[z_{t+l} - \hat{z}_t(l)]^2 = (1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2)\sigma_a^2 + \sum_{j=0}^{\infty} \left\{ \psi_{l+j} - \psi_{l+j}^* \right\}^2 \sigma_a^2 \qquad [8.2.8]$$

It can be seen that the above equation is minimized by setting $\psi_{l+j}^* = \psi_{l+j}$, $j = 0,1,2,\ldots,$ and thereby eliminating the second component on the right hand side of [8.2.8]. Consequently, when written in random shock form the MMSE forecast is derived as

$$\hat{z}_{t+l} = \psi_l a_t + \psi_{l+1} a_{t-1} + \psi_{l+2} a_{t-2} + \cdots \qquad [8.2.9]$$

As noted by Box and Jenkins (1976, Ch. 5) the finding in [8.2.9] is a special case of more general results in prediction theory by Wold (1954), Kolmogorov (1939, 1941a,b), Wiener (1949) and Whittle (1963). By combining the result in [8.2.9] with the random shock model in [8.2.6]

$$z_{t+l} = (a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \cdots + \psi_{l-1} a_{t+1})$$
$$+ (\psi_l a_t + \psi_{l+1} a_{t-1} + \psi_{l+2} a_{t-2} + \cdots)$$
$$= e_t(l) + \hat{z}_t(l) \qquad [8.2.10]$$

where $e_t(l)$ is the error of the MMSE forecast $\hat{z}_t(l)$.

8.2.3 Properties

The fact that [8.2.10] can be easily derived from the definition of a MMSE forecast, points out one of the advantages of using this kind of forecast. Fortunately, there are many other properties of a MMSE forecast that make it very beneficial for use in practical applications and some of them are described now. Let

$$E_t[z_{t+l}] = E[z_{t+l} | z_t, z_{t-1}, \cdots]$$

denote the *conditional expectation* of z_{t+l} given knowledge of all the observations up to time t. Then, attractive properties of a MMSE forecast include:

1. The MMSE forecast, $\hat{z}_t(l)$, is simply the conditional expectation of z_{t+l} at time t.

This can be verified by taking the conditional expectation of z_{t+l} in [8.2.10] to get

$$E_t[z_{t+l}] = \psi_l a_t + \psi_{l+1} a_{t-1} + \psi_{l+2} a_{t-2} + \cdots = \hat{z}_t(l) \qquad [8.2.11]$$

Keep in mind that when deriving [8.2.11] the expression $E_t[a_{t+k}] = 0$ for $k > 0$, and $E_t[a_{t+k}] = a_{t+k}$ for $k \leq 0$ since the innovations up to time t are known. Specific rules for calculating MMSE forecasts for any ARMA or ARIMA model are presented in the next subsection.

2. The *forecast error* is a simple expression for any ARMA or ARIMA model.

From [8.2.10], the forecast error from origin t and for lead time l is

$$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1} \qquad [8.2.12]$$

3. One can conveniently calculate the *forecast error variance*.

In particular, the variance of the forecast error is

$$E[e_t(l)^2] = V(l) = var[e_t(l)]$$
$$= E[(a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \cdots + \psi_{l-1} a_{t+1})^2]$$
$$= (1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2)\sigma_a^2 \qquad [8.2.13]$$

4. The MMSE forecast is *unbiased*.

This is because

$$E_t[e_t(l)] = E[a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \cdots + \psi_{l-1} a_{t+1}] = 0 \qquad [8.2.14]$$

where $E[a_{t+k}] = 0$ for $k > 0$.

5. The one step ahead forecast error is equal to the corresponding innovation and, therefore, one step ahead forecast errors are uncorrelated.

From [8.2.12], the one step ahead forecast error is

$$e_t(1) = z_{t+1} - \hat{z}_t(1) = a_{t+1} \qquad [8.2.15]$$

Because the innovations are independent, and, hence, uncorrelated, the one step ahead forecast errors must also be uncorrelated. As explained in Section 8.3, this result is very useful for developing tests to determine if one model forecasts significantly better than another.

6. Forecasts for lead times greater than one are, in general, correlated.

Consider forecasts for different lead times from the same origin t. Let the forecast errors for lead times l and $l+j$, where j is a positive integer, be given by $e_t(l)$ and $e_t(l+j)$, respectively. As shown by Box and Jenkins (1976, Appendix A5.1), the correlation between these two forecast errors is

$$corr[e_t(l), e_t(l+j)] = \frac{\sum_{i=0}^{l-1} \Psi_i \Psi_{j+i}}{\left\{ \sum_{h=0}^{l-1} \Psi_h^2 \sum_{g=0}^{l+j-1} \Psi_g^2 \right\}^2} \qquad [8.2.16]$$

Because of the correlation in [8.2.16], forecasts can lie either mainly above or below the actual observations when they become known.

7. Any linear function of the MMSE forecasts is also a MMSE forecast of the corresponding linear function of the future observations.

To explain what this means in practice, consider a simple example. Suppose that $\hat{z}_t(1)$, $\hat{z}_t(2)$, $\hat{z}_t(3)$ and $\hat{z}_t(4)$, are four MMSE forecasts. Then, $10\hat{z}_t(1) + 8\hat{z}_t(2) + 6\hat{z}_t(3) + 4\hat{z}_t(4)$ is a MMSE forecast of $10z_{t+1} + 8z_{t+2} + 6z_{t+3} + 4z_{t+4}$.

8.2.4 Calculation of Forecasts

Forecasting with ARMA Models

As explained in the previous subsection, the MMSE forecast, $\hat{z}_t(l)$, for lead time l, is simply the conditional expectation, $E_t[z_{t+l}]$, of z_{t+l} at origin t. When calculating the conditional expectations for an ARMA or ARIMA model one can write the model in any one of its three equivalent forms. These three formats are the difference equation form for the model as originally defined, random shock format and the inverted form (see Sections 3.4.3 and 4.3.4 for descriptions of the three forms for the ARMA and ARIMA family of models, respectively).

To simplify the notation required when determining MMSE forecasts, let the conditional expectations $E_t[a_{t+l}]$ and $E_t[z_{t+l}]$ be replaced by $[a_{t+l}]$ and $[z_{t+l}]$, respectively. For explaining how forecasts are determined using the three equivalent formats, consider the family of ARMA models defined in Chapter 3. For $l > 0$, the three equivalent formats for writing the MMSE forecasts are as follows:

Forecasts Using the Original Definition. By taking conditional expectations at time t of each term of the ARMA model in [3.4.3], the MMSE forecasts are:

$$[z_{t+l}] = \hat{z}_t(l) = \phi_1[z_{t+l-1}] + \phi_2[z_{t+l-2}] + \cdots + \phi_p[z_{t+l-p}] + [a_{t+l}]$$

$$- \theta_1[a_{t+l-1}] - \theta_2[a_{t+l-2}] - \cdots - \theta_q[a_{t+l-q}] \qquad [8.2.17]$$

As before, for convenience of explanation, the mean of the series is not written in the model. Following specific rules described below for calculating MMSE forecasts, one can easily determine each conditional expectation in [8.2.17].

Forecasts from the Random Shock Form. One can take conditional expectations at time t of the random shock form of the ARMA model in [3.4.18] to determine the MMSE forecasts as:

$$[z_{t+l}] = \hat{z}_t(l) = [a_{t+l}] + \psi_1[a_{t+l-1}] + \psi_2[a_{t+l-2}] + \cdots \qquad [8.2.18]$$

where ψ_i is the ith random shock parameter. When there are AR parameters in the original ARMA model, the number of innovation terms on the right hand side of [8.2.18] is infinite in extent. However, because the absolute values of the random shock parameters die off quickly for increasing lag, one can use a finite number of terms on the right hand side of [8.2.18] for calculating the forecasts up to any desired level of accuracy. Approaches for deciding upon how many MA parameters or terms to include in the random shock model are discussed in Section 3.4.3.

Forecasts using the Inverted Form. By taking conditional expectations at time t of the inverted form of the ARMA model in [3.4.25], the MMSE forecasts are:

$$[z_{t+l}] = \hat{z}_t(l) = [a_{t+l}] + \pi_1[z_{t+l-1}] + \pi_2[z_{t+l-2}] + \cdots \qquad [8.2.19]$$

where π_i is the ith inverted parameter. When there are MA parameters in the original model, the number of π_i parameters on the right side of [8.2.19] is infinite. Nonetheless, since the absolute values of the inverted parameters attenuate fairly quickly for increasing lag, only a finite number of inverted terms in [8.2.19] are required for calculating MMSE forecasts. Guidelines for deciding upon how many inverted components to include in the inverted form of the model are given in Section 3.4.3. In practice, only a moderate number of inverted parameters are needed.

Forecasting with an ARIMA Model

When forecasting with an ARIMA model, the simplest approach is to first calculate the *generalized nonseasonal AR operator* $\phi'(B)$ defined as

$$\phi'(B) = \phi(B)\nabla^d \qquad [8.2.20]$$

where $\phi(B)$ is the nonseasonal AR operator of order p, ∇^d is the nonseasonal differencing operator given in [4.3.3], and

$$\phi'(B) = 1 + \phi'_1B + \phi'_2B^2 + \cdots + \phi_{p+d}B^{p+d}$$

is the generalized nonseasonal AR operator for which ϕ'_i is the ith nonseasonal generalized AR parameter. The ARIMA model from [4.3.4] is then written as

$$\phi'(B)z_t = \theta(B)a_t \qquad [8.2.21]$$

where $\theta(B)$ is the nonseasonal MA operator of order q. By taking conditional expectations at time t of [8.2.21], the MMSE forecasts for an ARIMA model are determined using

$$[z_{t+l}] = \phi'_1[z_{t+l-1}] + \phi'_2[z_{t+l-2}] + \cdots + \phi'_{p+d}[z_{t+l-p-d}]$$
$$+ [a_{t+l}] - \theta_1[a_{t+l-1}] - \theta_2[a_{t+l-2}] - \cdots - [a_{t+l-q}] \qquad [8.2.22]$$

As pointed out in Section 4.3.1, usually the differenced series $w_t = \nabla^d z_t$ has a mean or level of zero. However, suppose this is not the case so that the model in [8.2.21] can be written as

$$\phi'(B)z_t = \theta_0 + \theta(B)a_t \qquad [8.2.23]$$

where the "deterministic component" $\theta_0 = \mu_w \phi'(1)$ and μ_w is the mean of the w_t series. For $d = 0$, 1 and 2 the term θ_0 can be interpreted as the level, slope in a linear deterministic trend, and quadratic trend coefficient, respectively. When the ARIMA model has the form of [8.2.23], forecasts are calculated recursively for $l = 1,2,\ldots,$ using

$$[z_{t+l}] = \theta_0 + \phi'_1[z_{t+l-1}] + \phi_2[z_{t+l-2}] + \cdots + \phi_{p+d}[z_{t+l-p-d}] + [a_{t+l}]$$
$$- \theta_1[a_{t+l-1}] - \theta_2[a_{t+l-2}] - \cdots - \theta_q[a_{t+l-q}] \qquad [8.2.24]$$

Rules for Forecasting

The most convenient equations to utilize when calculating MMSE forecasts are [8.2.17] and [8.2.22] for ARMA and ARIMA models, respectively. Whatever difference equation form of the ARMA or ARIMA model is employed for determining MMSE forecasts, one employs the simple rules listed below for the case of j being a non-negative integer to determine the conditional expectations written in these equations.

1. $$[z_{t-j}] = E_t[z_{t-j}] = z_{t-j}\,, \quad j = 0,1,2,\cdots \qquad [8.2.25]$$

   Because an observation at or before time t is known, the conditional expectation of this known value or constant is simply the observation itself.

2. $$[z_{t+j}] = E_t[z_{t+j}] = \hat{z}_t(j)\,, \quad j = 1,2,\cdots \qquad [8.2.26]$$

   The conditional expectation of a time series value after time t is the MMSE forecast that one wishes to calculate for lead time j from origin t.

3. $$[a_{t-j}] = E_t[a_{t-j}] = a_{t-j}\,, \quad j = 0,1,2,\cdots \qquad [8.2.27]$$

   Since an innovation at or before time t is known, the conditional expectation of this known value is the innovation itself. In practice, the innovations are not measured directly like the z_t's but are estimated when the ARMA or ARIMA model is fitted to the z_t or differenced series (see Chapter 6). Another way to determine a_t is to write [8.2.15] as

$$a_t = z_t - \hat{z}_{t-1}(1)$$

where $\hat{z}_{t-1}(1)$ is the one step ahead forecast from origin $t-1$.

4. $$[a_{t+j}] = E_t[a_{t+j}] = 0\,, \quad j = 1,2,\cdots \qquad [8.2.28]$$

In the definition of the ARMA or ARIMA model, the a_t's are assumed to be independently distributed and have a mean of zero and variance of σ_a^2. Consequently, the expected value of the unknown a_t's after time t is zero because they have not yet taken place.

8.2.5 Examples

To explain clearly how one employs the rules from the previous section for calculating MMSE forecasts for both ARMA and ARIMA, two simple illustrative examples are presented. The first forecasting application is for a stationary ARMA model while the second one is for a nonstationary ARIMA model.

ARMA Forecasting Illustration

ARMA(1,1) models are often identified for fitting to annual hydrological and other kinds of natural time series. For example, in Table 5.4.1, an ARMA(1,1) model is selected at the identification stage for fitting to an annual tree ring series.

From Section 3.4.1 an ARMA(1,1) model is written in its original difference equation form for time $t+l$ as

$$(1 - \phi_1 B)z_{t+l} = (1 - \theta_1 B)a_{t+l}$$

or

$$z_{t+l} - \phi_1 z_{t+l-1} = a_{t+l} - \theta_1 a_{t+l-1}$$

or

$$z_{t+l} = \phi_1 z_{t+l-1} + a_{t+l} - \theta_1 a_{t+l-1}$$

By taking conditional expectations of each term in the above equation, the ARMA(1,1) version for [8.2.17] is

$$[z_{t+l}] = \phi_1[z_{t+l-1}] + [a_{t+l}] - \theta_1[a_{t+l-1}] \qquad [8.2.29]$$

Using the rules listed in [8.2.25] to [8.2.28], one can calculate the MMSE forecasts for various lead times l from origin t.

Lead Time l=1:

Substitute $l = 1$ into [8.2.29] to get

$$[z_{t+1}] = \phi_1[z_t] + [a_{t+1}] - \theta_1[a_t]$$

After applying the forecasting rules, the one step ahead forecast is

$$\hat{z}_t(1) = \phi_1 z_t + 0 - \theta_1 a_t = \phi_1 z_t - \theta_1 a_t$$

In the above equation, all of the parameters and variable values on the right hand side are known, so one can determine $\hat{z}_t(1)$.

Lead Time l=2:

After substituting $l = 2$ into [8.2.29], one obtains

$$[z_{t+2}] = \phi_1[z_{t+1}] + [a_{t+2}] - \theta_1[a_{t+1}]$$

Next one uses the rules from [8.2.25] to [8.2.28] to get

$$\hat{z}_t(2) = \phi_1 \hat{z}_t(1) + 0 - \theta_1(0) = \phi_1 \hat{z}_t(1)$$

where the one step ahead forecast is known from the previous step for lead time $l = 1$.

Lead Time $l \geq 2$:

When the lead time is greater than one, the forecasting rules are applied to [8.2.29] to get

$$\hat{z}_t(l) = \phi_1 \hat{z}_t(l-1) + 0 - \theta_1(0) = \phi_1 \hat{z}_t(l-1)$$

where the MMSE forecast $\hat{z}_t(l-1)$ is obtained from the previous iteration for which the lead time is $l - 1$.

ARIMA Forecasting Application

In Section 4.3.3, the most appropriate ARIMA model to fit to the total annual electricity consumption for the U.S. is an ARIMA(0,2,1) model. From Figure 4.3.10, one can see that the series is highly nonstationary and, therefore, differencing is required.

Following the general form of the ARIMA model defined in [4.3.3] and [4.3.4], the ARIMA(0,2,1) model is written at time $t+l$ as

$$(1 - B)^2 z_{t+l} = (1 - \theta_1 B)a_{t+l}$$

or

$$(1 - 2B + B^2)z_{t+l} = (1 - \theta_1 B)a_{t+l}$$

or

$$z_{t+l} - 2z_{t+l-1} + z_{t+l-2} = a_{t+l} - \theta_1 a_{t+l-1}$$

After taking conditional expectations of each term in the above equation, the forecasting equation is

$$[z_{t+l}] - 2[z_{t+l-1}] + [z_{t+l-2}] = [a_{t+l}] - \theta_1[a_{t+l-1}]$$

or

$$[z_{t+l}] = 2[z_{t+l-1}] - [z_{t+l-2}] + [a_{t+l}] - \theta_1[a_{t+l-1}] \quad [8.2.30]$$

By employing the rules given in [8.2.25] to [8.2.30], one can determine the MMSE forecasts for lead times $l = 1,2, \cdots$, from origin t.

Lead Time $l = 1$:

Substitute $l = 1$ into [8.2.30] to obtain

$$[z_{t+1}] = 2[z_t] - [z_{t-1}] + [a_{t+1}] - \theta_1[a_t]$$

After invoking the forecasting rules, the one step ahead forecast is

$$\hat{z}_t(1) = 2z_t - z_{t-1} + 0 - \theta_1 a_t = 2z_t - z_{t-1} - \theta_1 a_t$$

Because all entries on the right hand side of the above equation are known, one can calculate $\hat{z}_t(1)$. Keep in mind that when fitting a model to a time series z_t, the historical z_t innovations are calculated at the estimation stage. Another way to calculate a_t is to write [8.2.15] as

$$a_t = z_t - \hat{z}_{t-1}(1)$$

where $\hat{z}_{t-1}(1)$ is the one step ahead forecast from origin $t-1$.

Lead Time $l = 2$:

After assigning $l = 2$ in [8.2.28], one gets

$$[z_{t+2}] = 2[z_{t+1}] - [z_t] + [a_{t+2}] - \theta_1[a_{t+1}]$$

In the next step, one uses the rules for calculating conditional expectations in order to obtain

$$\hat{z}_t(2) = 2\hat{z}_t(1) - z_t + 0 - \theta_1(0) = 2\hat{z}_t(1) - z_t$$

where the one step ahead forecast is determined in the previous iteration for which $l = 1$.

Lead Time $l = 3$:

Substitute $l = 3$ into [8.2.30] to obtain

$$[z_{t+3}] = 2[z_{t+2}] - [z_{t+1}] + [a_{t+3}] - \theta_1[a_{t+2}]$$

After applying the rules for calculating conditional expectations, the above equation becomes

$$\hat{z}_t(3) = 2\hat{z}_t(2) - \hat{z}_t(1) + 0 - \theta(0) = 2\hat{z}_t(2) - \hat{z}_t(1)$$

where the one and two step ahead forecasts from origin t are determined in the previous two iterations.

Lead Time $l \geq 3$:

When the lead time is greater than or equal to three, the forecasting rules are applied to [8.2.30] to obtain

$$\hat{z}_t(l) = 2\hat{z}_t(l-1) - \hat{z}_t(l-2) + 0 - \theta_1(0) = 2\hat{z}_t(l-1) - \hat{z}_t(l-2)$$

where the MMSE forecasts for $\hat{z}_t(l-1)$ and $\hat{z}_t(l-2)$ are determined in the two previous steps having lead times $l-1$ and $l-2$, respectively.

8.2.6 Updating Forecasts

When using the random shock form of the model, forecasts can be generated using [8.2.9] or [8.2.11]. The methods for calculating the random shock weights for ARMA and ARIMA models are presented in Sections 3.4.3 and 4.3.4, respectively. By using the random shock form of the forecasting model, one can develop an easy approach for efficiently updating forecasts. In particular, the forecasts $\hat{z}_{t+1}(l)$ and $\hat{z}_t(l+1)$ of the future observation z_{t+l+1} made from origins $t+1$ and t, respectively, are written following [8.2.11] as

$$\hat{z}_{t+1}(l) = \psi_l a_{t+1} + \psi_{l+1} a_t + \psi_{l+2} a_{t-1} + \cdots$$

$$\hat{z}_t(l+1) = \psi_{l+1} a_t + \psi_{l+2} a_{t-1} + \cdots$$

After subtracting the second equation from the first, one finds

$$\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + \psi_l a_{t+1} \qquad [8.2.31]$$

Because of this result, the forecast of z_{t+l+1} from origin t can be updated to become the forecast of z_{t+l+1} from origin $t+1$ by adding $\psi_l a_{t+1}$. From [8.2.15], one can see that a_{t+1} is simply the one step ahead forecast error from origin t.

In practice, the updating formula in [8.2.31] can be conveniently used for economizing on the number of computations for generating forecasts. Suppose one is at origin t and already has forecasts for lead times $l = 1,2,\ldots,L$. Immediately upon obtaining the next observation, z_{t+1}, one can calculate the forecast error $a_{t+1} = z_{t+1} - \hat{z}_t(1)$. This result can then be used to obtain forecasts $\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + \psi_l a_{t+1}$ from origin $t+1$ for lead times $l = 1,2,\ldots,L-1$. Although the new forecast $\hat{z}_{t+1}(L)$ cannot be calculated using this method, it can be easily determined from the forecasts at shorter lead times using the original difference equation form of the model (see Section 8.2.4).

8.2.7 Inverse Box-Cox Transformations

The overall procedure for determining forecasts from a time series model is displayed in Figure 8.1.1. Before fitting a model to a given series, one may wish to transform the series using the Box-Cox transformation in [3.4.30] or some other appropriate transformation. As explained in Section 3.4.5, the purpose of the transformation is to rectify problems with non-normality and/or heteroscedasticity in the residuals of the fitted model. Whatever the case, when one uses the model constructed for the transformed series to obtain MMSE forecasts following the methods of Section 8.2.4, one determines forecasts in the transformed domain. For example, when an ARMA model is built for a logarithmic average annual riverflow series, the forecasts from the model are MMSE forecasts of the logarithmic flows. As pointed out in Figure 8.1.1, to get forecasts in the untransformed domain, one must take some type of inverse transformation of the forecasts.

There are two basic approaches for determining forecasts in the original units of the series being forecasted. The first procedure is to take the direct inverse transformation of the forecasts produced in the transformed domain. For instance, suppose that the original z_t series is transformed using natural logarithms. From [3.4.30], this transformation is written as

$$z_t^{(\lambda)} = \ln(z_t + c)$$

where the constant c is chosen just large enough to cause all of the entries in z_t to be nonnegative. For an average annual riverflow series, c would be equal to zero. Using the techniques of Section 8.2.4, one can obtain MMSE forecasts for $z_t^{(\lambda)}$ from origin t for any desired lead times. To get the forecasts in the untransformed domain, one can use the direct inverse logarithmic transformation written as

$$\bar{z}_t(l) = \exp(\hat{z}_t^{(\lambda)}(l) - c) \qquad [8.2.32]$$

where $\hat{z}_t^{(\lambda)}(l)$, $l = 1,2,\ldots,$ is the MMSE forecast of $z_t^{(\lambda)}$ in the transformed domain and $\bar{z}_t(l)$ is the corresponding forecast in the untransformed domain. The symbol for a MMSE forecast is not written above the forecast in the untransformed domain because usually the direct inverse transformation of a MMSE forecast in the transformed domain does not produce a MMSE forecast in the untransformed domain. When not using logarithms, the direct inverse Box-Cox transformation of the MMSE forecasts in transformed domain is written in the untransformed format as

$$\bar{z}_t(l) = [\lambda \hat{z}_t(l) + 1]^{1/\lambda} - c \quad \text{where } \lambda \neq 0 \qquad [8.2.33]$$

Granger and Newbold (1976) call this the naive method since forecasts calculated using [8.2.37] or [8.2.38] are not the exact MMSE forecasts in the untransformed domain.

The second main approach for obtaining a forecast in the untransformed domain is to calculate the exact MMSE forecast (Granger and Newbold, 1976). More specifically, the exact MMSE forecast in the untransformed domain is determined from the fact that its transformed value follows a Normal distribution with expected value $\hat{z}_t^{(\lambda)}(l)$ and variance $V(l)$, where $V(l)$ is calculated using [8.2.13]. The expected value of the inverse Box-Cox transformed value is the desired MMSE forecast. Thus, the MMSE forecast, $\hat{z}_t(l)$, is given by

$$\hat{z}_t(l) = \frac{1}{\sqrt{2\pi V(l)}} \int_{-\infty}^{\infty} (\lambda y + 1)^{\frac{1}{\lambda}} e^{-\frac{1}{2}\frac{(y - \hat{z}_t^{(\lambda)}(l))^2}{V(l)}} dy, \quad \lambda \neq 0, \qquad [8.2.34]$$

and

$$\hat{z}_t(l) = e^{\hat{z}_t^{(\lambda)}(l) + \frac{1}{2}V(l)}, \quad \lambda = 0. \qquad [8.2.35]$$

The required integral in [8.2.34] may be determined numerically by Hermite polynomial integration.

In practice, it is found that the MMSE forecasts are slightly smaller than the corresponding naive forecasts. Also, studies with real data have shown that these minimum-mean-square-error forecasts do perform better than the naive forecasts.

8.2.8 Applications

Probability Limits

Models fitted to two annual time series are used for producing MMSE forecasts. In the first application, forecasts are calculated for an ARMA model describing a stationary series. The second forecasting example deals with forecasting using an ARIMA model fitted to a nonstationary series.

When plotting MMSE forecasts one should always include probability limits so that the variability in the forecasts can be properly appreciated. By using the formula for the variance of the forecast error in [8.2.13] and assuming normality one can calculate confidence limits. For example, the 50% probability limits for the l-step ahead MMSE forecast from origin t is

$$\hat{z}_t(l) \pm 0.674\sqrt{V(l)}$$

where $V(l)$ is the variance of the forecast error in [8.2.13]. When forecasting from origin t up to lead time L, one can calculate and plot the forecasts and 50% probability limits for $l = 1,2,\ldots,L$. Because the random shock parameters in [8.2.13] attenuate to zero for a stationary ARMA model, the forecasting probability limits asymptotically approach constant values for increasing l. On the other hand, the probability limits for forecasts from a nonstationary ARIMA model diverge for increasing l.

ARMA(1,1) Forecasts

A time series consisting of 700 tree ring indices from 1263 to 1962 is given by Stokes et al. (1973). The most appropriate ARMA model to fit to this series is the ARMA(1,1) model written in [3.4.15]. Following the rules given in Section 8.2.4, one can calculate MMSE forecasts for the calibrated tree ring model. Figure 8.2.2 displays the MMSE forecasts for lead times from 1 to 20. Notice that later observations in the series are plotted up to 1962. Starting from the origin 1962, MMSE forecasts are indicated from 1963 to 1982 along with their 50% and 90% probability intervals.

An example that explains how to calculate MMSE forecasts for an ARMA(1,1) model is given at the beginning of Section 8.2.5. Because the model is stationary, the forecasts for increasing lead times in Figure 8.2.2 draw closer to the mean of the series and the probability intervals run parallel to these forecasts. As would be expected, the best forecast for a future observation that is far from the last observation is the mean level.

ARIMA(0,2,1) Forecasts

Figure 4.3.10 portrays a graph of the total annual electricity consumption in the U.S.A. from 1920 to 1970 (United States Bureau of the Census, 1976). As explained in Section 4.3.3, the best ARIMA model to fit to this series is an ARIMA(0,2,1) model with an estimated Box-Cox transformation of $\lambda = 0.533$. Following the approach of Section 8.2.4, MMSE forecasts are first determined for the transformed domain where $\lambda = 0.533$. Subsequently, [8.2.35] is employed for calculating the MMSE forecasts shown in Figure 8.2.3 in the untransformed domain. An example of how to calculate MMSE forecasts by hand for an ARIMA(0,2,1) model without a Box-Cox transformation is given in Section 8.2.5.

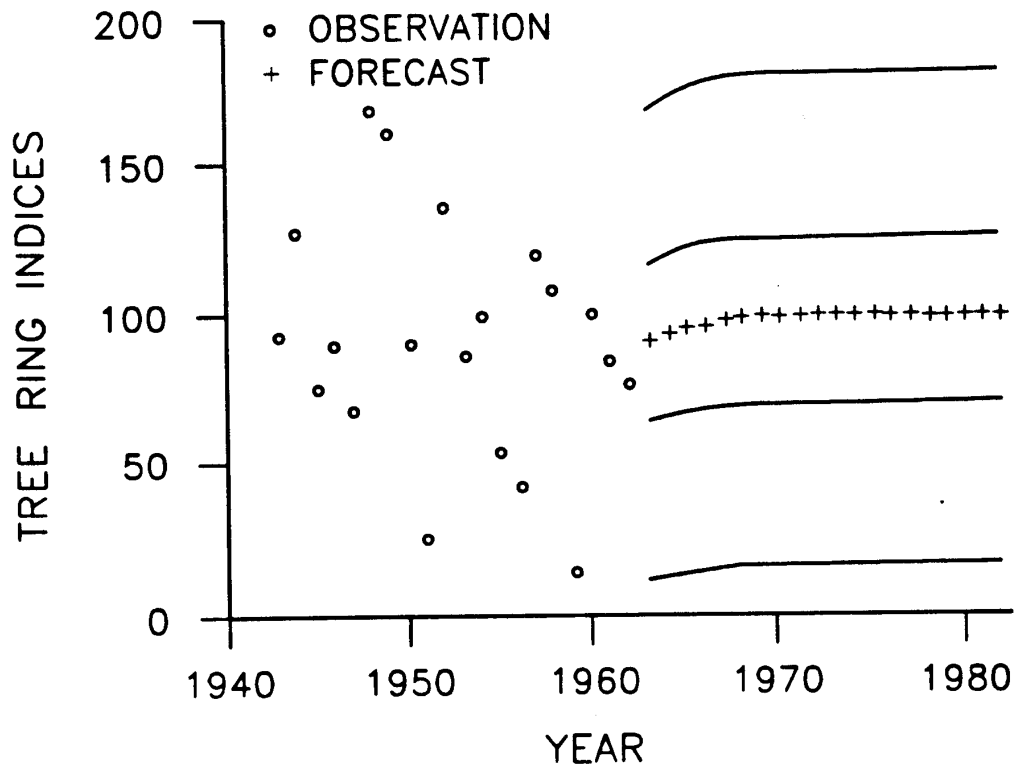

Figure 8.2.2. MMSE forecasts along with their 50% and 95% probability intervals for the ARMA(1,1) model fitted to the Douglas fir tree ring indices from Navajo National Monument, Arizona, U.S.A.

Figure 8.2.3 shows the MMSE forecasts in the untransformed domain calculated using the fitted model from 1971 to 1990 along with the 50% and 95% probability intervals. Because the series is nonstationary, observe how the forecasts continue the upward trend that is followed by the observations plotted on the left side of the figure. Moreover, the nonstationarity causes the probability limits to diverge outwards from the forecasts for increasing lead times.

8.3 FORECASTING EXPERIMENTS

8.3.1 Overview

An important test of the adequacy of a time series model is its ability to forecast well. The objective of this section is to employ forecasting experiments to demonstrate that ARMA models forecast very well when compared to other types of time series models that can be fitted to annual natural time series. This provides a sound reason for recommending the use of ARMA models by practitioners. In Sections 9.8 and 10.6, it is shown that ARMA models are also ideally suited for simulating hydrological as well as other types of natural phenomena.

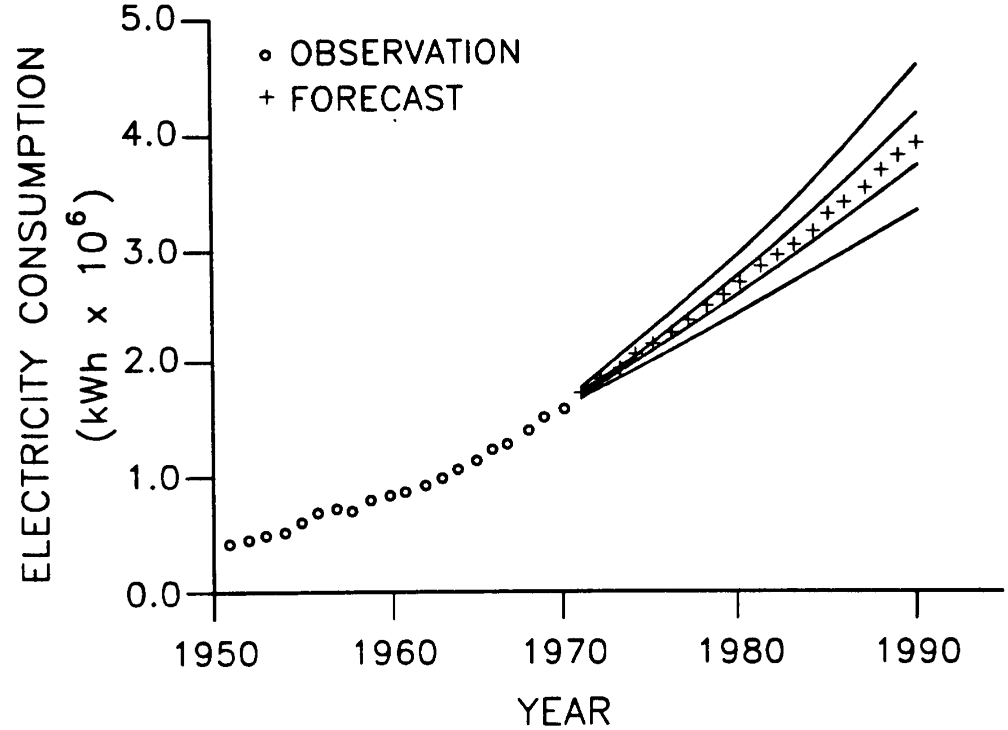

Figure 8.2.3. MMSE forecasts along with their 50% and 90% confidence intervals for the ARIMA(0,2,1) model fitted to the annual electricity consumption in the U.S.A.

In practical applications, *one step ahead forecasts* are often required for effectively operating a large-scale engineering project such as a system of reservoirs. When a new observation becomes available, the next one step ahead forecast can be made for deciding upon operating rules in the subsequent time period. Furthermore, a theoretical advantage of one step ahead forecasts is that they are statistically independent. This property allows one to develop statistical tests for determining if one model forecasts significantly better than another. In the next section, *statistical tests for comparing one step ahead forecasts* are presented and following this the different kinds of models used in the forecasting experiments are described.

To test the forecasting abilities of several stationary nonseasonal time series models, *split sample experiments* are performed in Section 8.3.4. Time series models are fitted to the first portion of the data in each of fourteen time series and these models are then employed to generate one step ahead forecasts. The forecasts errors are then compared using several loss functions to obtain ordinal rankings of the models. Statistical tests from Section 8.3.2 are then employed to test for significant differences in the forecasting performances of the various models. The forecasting results in the remaining part of Section 8.3 were originally presented by Noakes (1984) and Noakes et al. (1988).

8.3.2 Tests for Comparing Forecast Errors

Introduction

In the past, a great deal of effort has been devoted to the development of a wide variety of forecasting procedures. These procedures range from naive models or intuitive guesses to sophisticated techniques requiring skilled analysts and significant computer resources. At the same time, relatively little research has been devoted to developing methods for evaluating the relative accuracy of forecasts produced by the various forecasting methods.

In the forecasting experiments presented in Section 8.3.4, the forecast errors are examined from two different perspectives. Firstly, the performances of the various models are judged solely on the relative magnitudes of several criteria such as the ***mean squared error (MSE)*** or the ***mean absolute percentage error (MAPE)*** of the forecast errors. These comparisons provide ordinal rankings of the models but give no indication as to whether forecasts from a particular model are significantly better than forecasts from another model in a statistical sense. In order to address this question, a number of statistical tests are proposed to compare the performances of the models in a pairwise fashion and also to test the overall performances of particular models.

Wilcoxon Signed Rank Test

In order to ascertain whether the forecasts from a particular model are statistically significantly better than the forecasts generated by an alternative model, some form of statistical test must be employed. A nonparametric *Wilcoxon signed rank test* for paired data is one test which could be employed to test for significant differences in the forecasting ability of two procedures. This test was originally developed by Wilcoxon (1945) and is described in Appendix A23.2 in this book.

In this test, the differences in the squares of the forecast errors from two models for the same series are compared. These differences are ranked in ascending order, without regard to sign, and assigned ranks from one to the number of forecast errors available for comparison. The sum of the ranks of all positive differences is then computed as T in [A23.2.3] and compared to tabulated values in order to determine if the forecasts from a one model are significantly better than the forecasts from a competing model.

The results of this test may also be employed to examine the performances of the models across all of the series in the study. In this test, the probability associated with each T value is calculated by examining the area in the tail of the distribution. Fisher (1970, p. 99) presents a *combined level of significance test* such that

$$-2\sum_{i=1}^{k}\ln(p_i) \sim \chi^2_{2k} \qquad [8.3.1]$$

where p_i is the calculated probability associated with each T and k is the number of series considered in the test.

The Likelihood Ratio and Correlation Tests

It is of interest to examine statistically the difference in MSE's of the one step ahead predictor for two competing procedures in order to determine if the MSE's are significantly different. Thus, if $e_{1,t}$ and $e_{2,t}$ $(t = 1,2,\ldots,L)$ denote the L one step ahead forecast errors for models 1 and 2 respectively, the null hypothesis is

$$H_0: MSE(e_{1,t}) = MSE(e_{2,t}) \qquad [8.3.2]$$

where $MSE(e) = \langle e^2 \rangle$ and $\langle . \rangle$ denotes expectation. The alternative hypothesis, H_1, is the negation of H_0.

Granger and Newbold (1977, p. 281) have pointed out that a method originally developed by Pitman (1939) could be used to ascertain if one model forecasts significantly better than another. In this case, it is necessary to assume that $(e_{1,t}, e_{2,t})$ are jointly normally distributed with mean zero and are independent for successive values of t. In practice, the forecast errors may not be expected to satisfy all of the assumptions but these assumptions are probably a sensible first approximation. The assumptions of independence and zero mean seem quite reasonable if the forecasts are based on a good statistical model. As shown by Noakes (1984) and Noakes et al. (1988), a new test can be developed for the case in which the means are not known to be zero. For *Pitman's test*, let $S_t = e_{1,t} + e_{2,t}$ and $D_t = e_{1,t} - e_{2,t}$. Then Pitman's test is equivalent to testing if the correlation, r between S_t and D_t is significantly different from zero. Thus, provided $L > 25$, H_0 is significant at the five percent level if $|r| > 1.96/\sqrt{L}$. Previously, Pitman's test has often been used for testing the equality of variances of paired samples (Snedecor and Cochran, 1980, p. 190). It was pointed out in Lehmann (1959, p. 208, problem 33) that in this situation the test is unbiased and uniformly most powerful.

If the means of $e_{1,t}$ and $e_{2,t}$ are not both known to be zero, a likelihood ratio test can be employed. Let $(e_{1,t}, e_{2,t})$ be jointly normal with means (μ_1, μ_2) and covariance matrix

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

where σ_i^2 is the variance of the ith series and σ_{ij} is the covariance between i and j. Then the log likelihood for $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12})$ is given by Rao (1973, p. 448) as

$$\log L(.) = \frac{L}{2}\log|(\sigma^{ij})| - \frac{1}{2}\sum\sum\sigma^{ij}[S_{ij} + L(\mu_i - \bar{\mu}_i)(\mu_j - \bar{\mu}_j)] \qquad [8.3.3]$$

where

$$\bar{\mu}_i = \frac{1}{L}\sum_{t=1}^{L} e_{i,t}$$

and

$$S_{ij} = L\sum_{t=1}^{L}(e_{i,t} - \bar{\mu}_i)(e_{j,t} - \bar{\mu}_j)$$

and $(\sigma^{ij}) = (\sigma_{ij})^{-1}$.

If L_0 is the maximized log likelihood assuming the null hypothesis is true and L_1 is the maximized log likelihood assuming the alternative hypothesis is true, then the *likelihood ratio statistic* is given by

$$R = 2(L_1 - L_0) \quad [8.3.4]$$

When H_0 is true, it can be shown that $R \approx \chi_1^2$ (Rao, 1973).

If it is assumed that the means of the two error series are zero, then ignoring constants, the maximized log likelihoods are

$$L_1 = -\frac{L}{2}\log(\hat{\sigma}_1^2\hat{\sigma}_2^2 - \hat{\sigma}_{12}^2) \quad [8.3.5]$$

and

$$L_0 = -\frac{L}{2}\log(\hat{\sigma}^2 - \hat{\sigma}_{12}^2) \quad [8.3.6]$$

where $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the estimated forecast error variances for the two competing models, $\hat{\sigma}_{12}$ is the estimated covariance of the estimated forecast errors and

$$\hat{\sigma}^2 = \frac{\hat{\sigma}_1^2+\hat{\sigma}_2^2}{2} \quad [8.3.7]$$

The resulting likelihood ratio is then calculated using [8.3.4].

Equation [8.3.3] is easily maximized analytically when there are no restrictions on the parameters and so the maximized log likelihood is obtained. Under H_0

$$\sigma_1^2 + \mu_1^2 = \sigma_2^2 + \mu_2^2 \quad [8.3.8]$$

and the log likelihood may be maximized numerically over $(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\sigma_{12})$ with $\sigma_2^2 = \sigma_1^2 + \mu_1^2 - \mu_2^2$. The conjugate direction minimization algorithm of Powell (1964) with a penalty function to ensure that $\sigma_2^2>0$ is recommended. Thus, the likelihood ratio test statistic, R, which is χ_1^2 under H_0 is obtained from [8.3.4].

8.3.3 Forecasting Models

Introduction

Stationary nonseasonal time series models are of particular interest to hydrologists since they often wish to model annual time series that are approximately stationary over a specified time period and subsequently use the fitted models for forecasting and simulation. Furthermore, stationary nonseasonal models form the foundations for seasonal (see Part VI) models as well as other kinds of models (see Parts VII to IX).

When fitting a nonseasonal stationary model, or for that matter any type of stochastic model, to a given data set, one can follow the identification, estimation and diagnostic check stages of model construction described in Part III as well as elsewhere in the book. Figure III.1 depicts this systems design approach to model building while Figure 6.3.1 shows how the AIC can enhance model construction. All of the different kinds of models employed in the forecasting studies are carefully developed following this sensible approach to model building.

The five families of stationary nonseasonal models used in the study are as follows:

1. ARMA (see Chapter 3, Part III, and Section 8.2 for definition, model building and forecasting, respectively),
2. FGN (Fractional Gaussian Noise, see Section 10.4.5),
3. FARMA (fractional ARMA, see Chapter 11) and FDIFF (fractioning differencing, special case of FARMA models in Chapter 11),
4. Markov (see this section),
5. Nonparametric (see this section).

Following the definition in Section 2.5.3, the second and third models have *long memory* while the remaining ones possess *short memory*. Additionally, the first three types of models are described at other indicated sections in this book while the last two are now outlined.

Markov and Nonparametric Regression Models

A number of researchers have proposed various nonparametric models for modelling and forecasting hydrological time series (see for example Denny et al. (1974) and Yakowitz (1973, 1976, 1979a,b, 1985a,b)). These models offer an attractive alternative to the ARMA as well as long memory FGN and FARMA models. The flexibility and modest computational requirements associated with nonparametric models are certainly two important considerations in model selection. As well, probability statements can be made concerning forecasted events. In light of these attractive characteristics, two nonparametric models are considered in this forecasting study.

A First Order Markov Model: The underlying concepts associated with stationary Markov chains are well known and explained in many standard statistical and operational research books. The first model considered is a first order Markov process defined as

$$Pr(X_{n+1} | X_n, X_{n-1}, \cdots) = Pr(X_{n+1} | X_n) \qquad [8.3.9]$$

Although higher order processes may be required to adequately model the data, the first order approximation is a reasonable first step.

The time series data are first arranged in ascending order. If there are n data, m = integer $(\sqrt{n})$ states are selected at equal intervals. For example, if $n = 100$, then 10 states would be selected. The first 10 data would then be assigned to the first state and the state mean would be the arithmetic mean of these elements. This procedure is repeated until the m state means are calculated.

Based upon this arbitrary selection of states and estimated state means, each datum is reassigned to a specific state according to the Euclidean distance between the observation and the state means. That is, X_i is in state v if

$$|X_i - c_v| \le |X_i - c_k|, \quad 1 \le k \le m \qquad [8.3.10]$$

where the c_k's are the state means. A check is then made to ensure that at least $n^{1/3}$ data are associated with each state.

Quasi state transition probabilities are then estimated using the original time series and the selected states. Forecasts can then be made using these transition probabilities and the state means.

A Nonparametric Model: Yakowitz (1985a,b) employs nonparametric regression techniques to develop a more comprehensive and flexible nonparametric model. Unlike the simple first order Markov model outlined above, this nonparametric model allows for higher order dependence. A method for forecasting using this new model is also presented by Yakowitz (1985a,b).

Kernel nonparametric estimators of the density by Rosenblatt (1956, 1971) as well as kernel nonparametric regression estimators introduced by Watson (1964) have been extensively investigated and have also found practical application in fields such as pattern recognition. They can be briefly described as follows. Suppose that there are n independent observations, Y_i, $i = 1,2, \ldots, n$ with common density $f(y)$. Then the estimate of $f(y)$ based on the kernel $k(\cdot)$ is given by

$$\hat{f}(y) = \frac{1}{n\alpha_n} \sum_{i=1}^{n} k\left(\frac{y - Y_i}{\alpha_n}\right) \qquad [8.3.11]$$

where α_n is called a smoothing parameter and $k(\cdot)$ is generally taken to be a probability density function such as the standard normal. The choice of the kernel, $k(\cdot)$, is not as crucial as is the choice of the parameter α_n to obtain a good estimate.

For the regression case, suppose that one observes pairs of independent and identically distributed variables (Y_i,X_i) and that one wishes to estimate the expectation of $g(Y)$ conditional on the value $X = x$, where the pair (Y,X) has the same distribution as the observations (Y_i,X_i), $i = 1,2, \ldots, n$, and $g(\cdot)$ is a real function. The estimate of $E[g(Y)|X = x]$ is given by (Watson, 1964)

$$\hat{E}[g(Y)|X = x] = \frac{\sum_{i=1}^{n} g(Y_i) k\left(\frac{x - X_i}{\alpha_n}\right)}{\sum_{i=1}^{n} k\left(\frac{x - X_i}{\alpha_n}\right)} \qquad [8.3.12]$$

The extension of these estimators to the case where the observations form a dependent but stationary sequence has been accomplished by several authors (see, for example, Yakowitz (1985a,b), Collomb (1983, 1984), and Bosg (1983)). Suppose that Y_t is a time series process. Then [8.3.11] is an estimate of the marginal density function and if $X_i = Y_{i-1}$ then [8.3.12] is an estimate of $E[g(Y_t)|Y_{t-1} = y]$. The main condition for the use of the estimators [8.3.11] and [8.3.12] when Y_t is a stationary process is that they satisfy some kind of asymptotic independence such as geometric ergodicity (Yakowitz, 1985a). Note that if the process is Markov, $E[g(Y_t|Y_{t-1} = y)]$ is the optimal estimate of $g(Y_t)$ given the whole past under a least squares criterion. The main advantage of the estimators is the great flexibility that they provide to model nonlinearities when the nature of the departure from linearity is not obvious, as is the case in hydrological time series.

The higher order extensions of [8.3.11] and [8.3.12] are obvious and, hence, are not presented here. The choice of the parameter, α_n, is critical to obtain a balance between reduction of bias and reduction of variance of the estimates. The following procedure is employed to determine α_n for the models. For each point in the training set, one estimates the conditional regression function based on the rest of the training samples and obtains the sum of squares of the difference between the observed value and the estimate. This procedure is repeated for a range of values for α_n within which the absolute minimum of the sum of squares is found. The value of α_n which yields the minimum sum of squares is selected.

8.3.4 Forecasting Study

Introduction

To compare the forecasting performance of the various nonseasonal models mentioned in Section 8.3.3, two split sample experiments are performed. Annual river flow, tree ring indices, mud varve and annual temperature series are considered in these studies. Nonseasonal models are fitted to the first parts of the series and these models are then employed to forecast the remaining data.

Forecasting can, in fact, be used as a means of model discrimination among competing models. For a given type of data such as hydrological time series, select the class of models which forecast the best according to certain criteria. In economics, authors such as Granger and Newbold (1977) and Makridakis and Hibon (1979) have carried out extensive forecasting experiments to determine the best kinds of models to use with nonseasonal and seasonal data. Although water resources engineers have recognized the importance of forecasting for a long time, very few large forecasting studies have been executed. Consequently, the forecasting study presented in this section as well as by Noakes et al. (1988) and Noakes (1984) constitutes one of the first extensive forecasting studies in water resources. Forecasting experiments with seasonal and transfer function-noise models are given in Chapters 15 and 18, respectively.

A comprehensive approach for carrying out forecasting experiments is depicted in Figure 8.3.1. In the forecasting study reported here none of the series are first transformed before fitting the five models listed in Section 8.3.3 to the first part of the series. Furthermore, when forecasting the last part of the series, one step ahead forecasts are determined. As shown below both the ARMA and nonparametric regression model of Yakowitz forecast better than the other three kinds of models listed in the previous section. Finally, as demonstrated by the simulation experiments carried out in Section 10.6, ARMA models are capable of statistically preserving important historical statistics of annual geophysical time series.

First Forecasting Experiment

The annual data sets considered in the first study are listed in Table 8.3.1. The riverflow and temperature data are obtained from Yevjevich (1963) and Manley (1953), respectively. The most appropriate type of ARMA models to fit to the last two series in Table 8.3.1 are given in Table 5.4.1.

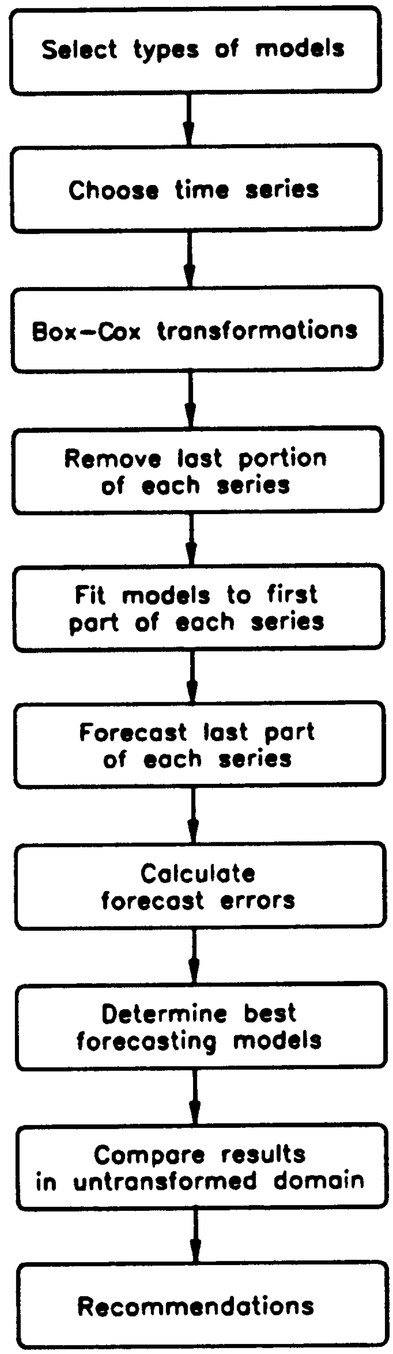

Figure 8.3.1. Forecasting experiments.

Because of the computational effort required to forecast using the FGN and FDIFF models, only series with less than 150 data are considered in the first study. The general procedure is to truncate the data sets by omitting the last 30 years of data. Models are then calibrated using the first portion of the data. These models are then employed to forecast one step ahead MMSE forecasts (see Section 8.2) of the last 30 years of data. For a given model and time series, one can calculate the forecasting error for each of the 30 one step ahead forecasts. By summing the squared forecast errors, dividing by 30 and then taking the square root of this, one obtains the *root mean square error (RMSE)* for the forecasts.

The RMSE's for the 30 one step ahead MMSE forecasts for each of the models entertained are given in Table 8.3.2. A summary of these results is presented in Table 8.3.3. The rank sum is simply the sum of the product of the rank and the associated table entry. Thus, models with low rank sums forecast better overall than models with higher rank sums. In this study, the nonparametric model proposed by Yakowitz (1985a,b) forecasts well for the time series considered while the FDIFF model is the worst model.

Pitman's test (see Section 8.3.2) is employed to test for statistically significant differences in the RMSE's of the forecasts. The five competing procedures are compared in a pairwise fashion. The correlation values, r, are presented in Table 8.3.4. For these r values, the 95% confidence limits are calculated to be $\pm \frac{1.96}{\sqrt{30}} = \pm 0.358$. The ARMA, Markov, FGN and nonparametric forecasts are all significantly better at the 5% significance level of $\pm$ 0.358 than the FDIFF forecasts for the series Ogden. The nonparametric forecasts are also significantly (0.05 level) better than the FGN forecasts for the series Ogden.

Second Forecasting Experiment

The data sets employed in the second study are listed in Table 8.3.5. Except for the Snake time series, the tree ring indices are from Stokes et al. (1973). The Snake tree ring indices are from Schulman (1956). The most appropriate type of ARMA model to fit to the Navajo series is listed in Table 5.4.1 as being ARMA(1,1).

The RMSE's of the ARMA, Markov, nonparametric and FARMA forecasts are presented in Table 8.3.6 while a summary of these results is given in Table 8.3.7. In all cases, the Markov model has the largest RMSE of the four models considered in this study. The ARMA and nonparametric models are essentially equal in performance and are both slightly better than the FARMA model.

The likelihood ratio test described in Section 8.3.2 is employed to test for significant differences between the ARMA and Markov forecast errors. In this case, the test statistic, R, is calculated for both instances, where the means of the forecast errors are assumed to be zero ($R1$) and non-zero ($R2$). The calculated values are presented in Table 8.3.8. There is virtually no difference between $R1$ and $R2$ so either value may be employed in the test. In this study, the ARMA forecasts are significantly (0.05 level) better than the Markov forecasts for the two series Eaglecol and Lakeview. Since the RMSE's of the ARMA models are always less than the RMSE's of the Markov models, the Markov forecasts could never be significantly better than the ARMA forecasts.

Table 8.3.1. Annual riverflow and temperature data sets.

Code names	River or data types	Locations	Periods	n
Gota	Gota	Sjotorp-Vanersburg, Sweden	1807-1957	150
Mstouis	Mississippi	St. Louis, Missouri	1861-1957	96
Neumunas	Neumunas	Smalininkai, USSR	1811-1943	132
Ogden	St. Lawrence	Ogdensburg, New York	1860-1957	97
Temp	Temperature	Central England	1802-1951	150

Table 8.3.2. RMSE's for the one step ahead forecasts for the annual riverflow and temperature series.

Code names	ARMA	FGN	FDIFF	Markov	Nonparametric
Gota	87.58	95.57	97.66	97.45	92.86
Mstouis	1508.03	1543.56	1574.85	1625.90	1560.00
Neumunas	118.30	115.80	116.12	114.70	115.40
Ogden	473.89	630.55	875.91	450.85	426.90
Temp	1.21	1.17[a]	1.17	1.13	0.95

[a] Indicates smaller of tied values.

Table 8.3.3. Distribution of the RMSE's for 30 forecasts for the annual riverflow and temperature series.

Ranks	Number of times each model has indicated rank				
	ARMA	FGN	FDIFF	Markov	Nonparametric
1	2	0	0	1	2
2	0	1	0	2	2
3	1	3	0	0	1
4	0	1	3	1	0
5	2	0	2	1	0
Rank sum	15	15	22	14	9

Pitman's test is employed to compare the ARMA, nonparametric and FARMA forecasts in a pairwise fashion. The calculated correlations, r, between S_t and D_t are presented in Table 8.3.9. The only significant value (0.05 level) is for the series Lakeview when the ARMA and nonparametric forecasts are compared. Thus, the ARMA forecasts are significantly better than the nonparametric forecasts for this series at the 5% level. In all other cases, there is no statistically significant difference in the forecasts produced by the various models.

Table 8.3.4. Pitman's correlations, r, for pairwise comparisons of 5 annual models for each of the 5 series.

	Gota	Mstouis	Neumunas	Ogden	Temp
A vs B	-0.170	-0.112	0.125	-0.347	0.112
A vs C	-0.223	-0.171	0.089	-0.593	0.103
A vs D	-0.302	-0.317	0.102	0.076	0.165
A vs E	-0.277	-0.193	0.142	0.160	0.241
B vs C	-0.142	-0.275	-0.049	-0.828	-0.092
B vs D	-0.060	-0.209	0.040	0.335	0.142
B vs E	0.063	-0.123	0.041	0.453	0.209
C vs D	0.008	-0.114	0.053	0.582	0.143
C vs E	0.123	0.083	0.096	0.663	0.212
D vs E	0.178	0.167	-0.029	0.081	0.180

*Models: ARMA = A, FGN = B, FDIFF = C, Markov = D, Nonparametric = E.

Discussion

Based upon the result of the forecasting studies, the use of FGN and FDIFF models for forecasting annual hydrological and tree ring time series is not recommended. The two models which should be given serious consideration are the nonseasonal ARMA model and the nonparametric model presented by Yakowitz (1985a). Both forecast equally well for the series considered in the studies presented in this section. Moreover, Noakes (1989) demonstrates that the nonparametric model works well for generating inseason forecasts of salmon returns.

The performance of the various models is evaluated using the RMSE's of the forecasts and some of the statistical tests outlined in Section 8.3.2. This assumes that identical costs are assigned to both negative and positive forecast errors of the same magnitude. One recognizes that an asymmetric loss function may be more appropriate in certain instances, particularly in hydrological applications. For instance, different costs may be associated with inaccurate forecasts that result in either a flood or a drought. However, the RMSE criterion is employed since the procedures used for estimating the model parameters involve minimizing the sum of squared error terms. Presumably, if the type of loss function to be used to evaluate the forecast performance is known a priori, then the parameter estimation procedures could be adapted to minimize the expected loss. Without prior knowledge of the type of loss function, the RMSE criterion would appear to be a reasonable compromise (Noakes et al., 1985, 1988).

8.4 CONCLUSIONS

By following the model construction procedure of Part III, one can develop a parsimonious ARMA or other type of model for describing a given time series. As explained in Section 8.2, one can then use this model to produce MMSE forecasts of future observations. If one wishes to compare the forecasting accuracy of a range of models for a specified kind of time series, one can use the general model discrimination procedure outlined in Figure 8.3.1. By using tests from Section 8.3.2, one can ascertain if one model forecasts one step ahead forecasts significantly better than another. The results of the forecasting experiments of Section 8.3 demonstrate that ARMA models forecast annual hydrological and tree ring series just as well or better than any of

Table 8.3.5. Tree ring indices data.

Code names	Types of Trees	Locations	Periods	n
Bigcone	Bigcone spruce	Southern California	1458-1966	509
Dell	Limber pine	Dell, Montana	1311-1965	655
Eaglecol	Douglas fir	Eagle, Colorado	1107-1964	858
Exshaw	Douglas fir	Exshaw, Alberta	1460-1965	506
Lakeview	Ponderosa pine	Lakeview, Oregon	1421-1964	544
Naramata	Ponderosa pine	Naramata, B.C.	1415-1965	515
Navajo	Douglas fir	Navajo National Monument, Belatakin, Arizona	1263-1962	700
Ninemile	Douglas fir	Ninemile Canyon, Utah	1194-1964	771
Snake	Douglas fir	Snake River Basin	1282-1950	669

Table 8.3.6. RMSE's of the last half of the tree ring series forecasted.

Code names	ARMA	Markov	Nonparametric	FARMA
Bigcone	38.52	39.01	38.33	38.83
Dell	36.83	37.73	37.41	37.16
Eaglecol	27.73	29.00	28.11	27.60
Exshaw	32.70	33.58	32.51	32.77
Lakeview	16.75	17.78	17.11	16.86
Naramata	29.98	30.75	30.16	30.18
Navajo	44.27	44.46	44.17	44.39
Ninemile	38.18	38.53	37.93	37.78
Snake	21.87	22.43	21.74	21.78

Table 8.3.7. Distribution of the RMSE's for the ARMA, Markov, Nonparametric and FARMA models when the last half of the tree ring series forecasted.

Ranks	Number of times each model has indicated rank			
	ARMA	Markov	Nonparametric	FARMA
1	3	0	4	2
2	4	0	2	3
3	2	0	3	4
4	0	9	0	0
Rank sum	17	36	17	20

Table 8.3.8. ARMA vs Markov likelihood ratio statistics for the last half of the tree ring series forecasted.

Code names	$R1$[a]	$R2$[b]
Bigcone	0.587	0.587
Dell	2.160	2.157
Eaglecol	6.667	6.665
Exshaw	3.036	3.032
Lakeview	9.323	9.324
Naramata	2.056	2.053
Navajo	0.176	0.176
Ninemile	0.694	0.691
Snake	2.381	2.381

[a] The means of the forecast errors are assumed to be zero.

[b] The means of the forecast errors are not assumed to be zero.

Table 8.3.9. Pairwise comparison of the ARMA, Nonparametric and FARMA models using Pitman's test and forecasting the last half of the tree ring series.

Code names	ARMA vs Nonparametric	ARMA vs FARMA	FARMA vs Nonparametric
Bigcone	-3.79E-2	-6.49E-3	-7.52E-2
Dell	8.83E-2	2.92E-4	8.76E-2
Eaglecol	3.43E-2	4.37E-2	4.41E-2
Exshaw	-7.95E-2	-1.57E-2	-1.57E-2
Lakeview	1.21E-1[a]	-8.88E-3	5.61E-2
Naramata	7.94E-2	-1.09E-2	-7.25E-2
Navajo	-1.83E-2	-2.14E-3	-2.93E-2
Ninemile	-3.89E-2	1.83E-2	2.23E-2
Snake	-3.55E-2	5.36E-3	1.68E-3

[a] Significant at the 5% level.

its competitors. For this and many other reasons, ARMA models are highly recommended for use in practical applications.

When forecasting, it is important to use models that provide an adequate fit to the data using as few model parameters as possible. For certain types of models, Ledolter and Abraham (1981) demonstrate that if a nonparsimonious model is employed for forecasting, the variance of the forecast errors increases. This problem may not be serious for large samples but for a small number of observations the effect of overfitting may not be negligible.

When using the techniques of Section 8.2 to calculate MMSE forecasts, one assumes that the model parameters are known exactly. However, in practice one must estimate the model parameters from the data. The uncertainty contained in the parameter estimates could be considered when forecasting. Several authors (Akaike, 1970; Bloomfield, 1972; Bhansali, 1974; Box and Jenkins, 1976, p. 267; Baillie, 1979) present results for the variance of the forecast error when fitted parameters are used in various time series models under the unrealistic assumption that a forecasted data point is independent of the data employed for parameter estimation. Extending earlier work of Phillips (1979), Kheoh (1986) plus Kheoh and McLeod (1989) develop an expression for the l-step ahead forecast error of an AR(1) model for which the effect on the variance of the forecast error when the parameter is estimated from the same data upon which the forecast is based is taken into account. In particular, the effect of estimating the parameter is to cause a reduction in the variance of the forecast error.

Forecasting procedures similar to those developed for nonseasonal models, can also be extended for use with seasonal and other types of models. In Chapter 15, for example, MMSE forecasts are calculated and compared for three different types of seasonal models. Forecasting results for transfer-function noise models having one output series and multiple inputs, are presented in Chapter 18.

When one can select a range of models to fit to a time series, one may wish to select the model that forecasts most accurately. An alternative approach is to combine the forecasts from two or more models in accordance with their relative performances. In this way, one may be able to take advantage of the forecasting strengths of each of the models. In Section 15.5.2, specific techniques are developed for combining forecasts in an optimal manner from various models in order to attempt to improve the overall accuracy of the resulting forecasts. In addition, two case studies are presented in Sections 15.5.2 and 18.4.2 for examining the utility of combining forecasts. Similar combination techniques could, of course, be used with nonseasonal models.

PROBLEMS

8.1 For an ARMA(2,1) model, calculate MMSE forecasts up to a lead time of 10.

8.2 Determine MMSE forecasts for an ARIMA(1,2,1) model up to a lead time of 10.

8.3 As explained in Section 4.3.3, the most appropriate model to fit to the U.S. electrical demand series is an ARIMA(0,2,1) model with $\lambda = 0.5$. The last two data points in the transformed series are $z_t^{(\lambda)} = 2561$ and $z_{t-1} = 2491$ while $\hat{a}_t = -13.32$, $\sigma_a^2 = 636.7$ and $\hat{\theta}_1 = 0.9563$. Calculate by hand the MMSE forecasts along with their 50% probability limits up to five steps ahead from the last observation. Compare your results to Figure 8.2.3.

8.4 Suppose that an ARIMA(0,1,1) model is written as

$$\nabla z_t = a_t - 0.4a_{t-1}$$

Generate forecasts for lead times $l = 1,2,3$ from origin t when the model is written as given above, in random shock form and inverted form.

8.5 Suppose that the best model to fit to a time series represented by z_t, $t = 1,2, \ldots, n$ is

$$(1 - B)\ln z_t = (1 - \theta_1 B)a_t$$

From origin time t calculate the MMSE forecasts for $l = 1,2, \ldots, 6$ in both the transformed and untransformed domains.

8.6 Fit the most appropriate ARMA or ARIMA model to a series of your choice. Plot forecasts along with 50% probability limits from the last data point up to lead time $l = 20$.

8.7 Take a yearly series having at least 70 observations and fit an ARMA or ARIMA model to the reduced version of this series that omits the last 20 data points. Using the calibrated model, calculate one step ahead forecasts and plot them against the known observations. Using this graph and other appropriate calculations, comment upon how well your model forecasts.

8.8 On December 13, 1978, Makridakis and Hibon (1979) read their paper on an empirical investigation of forecasting accuracy before the Royal Statistical Society in London. Summarize some of the main empirical findings of these authors. At the end of Makridakis and Hibons' paper, comments by attendees are presented. Mention some of the more interesting criticisms made about their paper and comment upon how well the authors defended themselves.

8.9 Explain how parameter uncertainty can be considered when forecasting with ARMA models.

8.10 Why can a nonparsimonious model increase the variance of the forecast errors of MMSE forecasts generated by an ARMA model?

REFERENCES

DATA SETS

Manley, G. (1953). The mean temperatures of Central England (1698-1952). *Quarterly Journal of the Royal Meteorological Society*, 79:242-261.

Schulman, E. (1956). *Dendroclimatic Changes in Semi-Arid America*. University of Arizona Press, Tucson, Arizona.

Stokes, M. A., Drew, L. G., and Stockton, C. W. (1973). Tree ring chronologies of Western America. Chronology Series 1, Laboratory of Tree Ring Research, University of Arizona, Tucson, Arizona.

United States Bureau of the Census (1976). The statistical history of the United States from colonial times to the present.

Yevjevich, V. M. (1963). Fluctuation of wet and dry years, 1, Research data assembly and mathematical models. Hydrology paper no. 1, Colorado State University, Fort Collins, Colorado.

FORECASTING

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

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

Granger, C. W. J. and Newbold, P. (1977). *Forecasting Economic Time Series.* Academic Press, New York.

Kolmogorov, A. N. (1939). Sur l'interpolation et l'extrapolation des suites stationnaires. *C. R. Acad. Sci. Paris,* 208:2043-2045.

Kolmogorov, A. N. (1941a). Interpolation und extrapolation von stationaren zufalligen folgen. *Bull. Acad. Sci. (Nauk) U. R. S. S. Ser. Math.,* 5:3-14.

Kolmogorov, A. N. (1941b). Stationary sequences in Hilbert space. *(in Russian). Bull. Math. Univ. Moscow,* 2(6):1-40.

Ledolter, J. and Abraham, B. (1981). Parsimony and its importance in time series forecasting. *Technometrics,* 23(4):411-414.

Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting,* 1:111-153.

Makridakis, S. and Hibon, M. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society, Series A,* 142:97-145.

Newbold, P. and Granger, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts, *Journal of the Royal Statistical Society,* Series A, 137:131-165.

Whittle, P. (1963). Prediction and Regulation. *English Universities Press,* London.

Wiener, N. (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time Series.* Wiley, New York.

Wold, H. O. (1954). *A Study in the Analysis of Stationary Time Series.* Almquist and Wicksell, Uppsala, Sweden, second edition.

HYDROLOGICAL FORECASTING

International Association of Hydrological Sciences (1980). Hydrological Forecasting Symposium, *Proceedings of the Oxford Symposium,* held April 15-18, 1980, IAHS-AISH Publication No. 129.

Noakes, D. J. (1984). Applied Time Series Modelling and Forecasting. PhD thesis, Dept. of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada.

Noakes, D. J. (1989). A nonparametric approach to generating inseason forecasts of salmon returns. *Canadian Journal of Fisheries and Aquatic Sciences,* 46(10):2046-2055.

Noakes, D. J., Hipel, K. W., McLeod, A. I., Jimenez, J. and Yakowitz, S. (1988). Forecasting annual geophysical time series. *International Journal of Forecasting,* 4:103-115.

Noakes, D. J., McLeod, A. I., and Hipel, K. W. (1985). Forecasting monthly riverflow time series. *International Journal of Forecasting*, 1:179-190.

NONPARAMETRIC MODELLING

Bosg, D. (1983). Nonparametric prediction in stationary processes. In *Lecture Notes in Statistics*, volume 16. Springer-Verlag, New York.

Collomb, G. (1983). From parametric regression to nonparametric prediction: Survey of mean square error and original results on the predictogram. In *Lecture Notes in Statistics*, volume 16. Springer-Verlag, New York.

Collomb, G. (1984). Proprietes de convergence presque complete du predictive a noyau. *Zeitschrift fur Wahrschein-lichkeitstheorie und Verwandte Gebiete*, 66:441-460.

Denny, J., Kisiel, C. and Yakowitz, S. J. (1974). Procedures for determining the order of Markov dependence in streamflow records. *Water Resources Research*, 10(5):947-954.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of density functions. *Annals of Mathematical Statistics*, 27:832-837.

Rosenblatt, M. (1971). Curve estimates. *Annals of Mathematical Statistics*, 42:1815-1842.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya*, Series A, 26:359-372.

Yakowitz, S. J. (1973). A stochastic model for daily river flows in arid regions. *Water Resources Research*, 9(5):1271-1285.

Yakowitz, S. J. (1976). Small-sample hypothesis tests of Markov order, with application to simulated hydrologic chains. *Journal of the American Statistical Association*, 71(353):132-136.

Yakowitz, S. J. (1979a). A nonparametric Markov model for daily river flow. *Water Resources Research*, 15(5):1035-1043.

Yakowitz, S. J. (1979b). Nonparametric estimation of Markov transition functions. *The Annals of Statistics*, 7:671-679.

Yakowitz, S. J. (1985a). Nonparametric density estimation and prediction for Markov sequences. *Journal of the American Statistical Association*, 80:215-221.

Yakowitz, S. J. (1985b). Markov flow models and the flood warning problem. *Water Resources Research*, 21(1):81-88.

PARAMETER UNCERTAINTY IN FORECASTING

Akaike, H. (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics*, 22:203-217.

Baillie, R. T. (1979). Asymptotic prediction mean squared error for vector autoregressive models. *Biometrika*, 66:675-678.

Bhansali, R. J. (1974). Asymptotic mean-square error of predicting more than one-step ahead using the regression model. *Applied Statistics*, 23:35-42.

Bloomfield, P. (1972). On the error prediction of a time series. *Biometrika*, 59:501-507.

Kheoh, T. S. (1986). Topics in Time Series Analysis and Forecasting. PhD thesis, Dept. of Statistical and Actuarial Sciences, The University of Western Ontario, London, Ontario, Canada.

Kheoh, T. S. and McLeod, A. I. (1989). On the forecast error variance of fitted time series models. *Pakistan Journal of Statistics*, 5(3):239-243.

Phillips, P. C. B. (1979). The sampling distribution of forecasts from a first-order autoregression. *Journal of Econometrics*, 9:241-261.

STATISTICAL TESTING

Fisher, R. A. (1970). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburg, England.

Lehmann, E. L. (1959). *Testing Statistical Hypothesis*. Wiley, New York.

Pitman, E. J. G. (1939). A note on normal correlation. *Biometrika*, 31:9-12.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. John Wiley, New York, second edition.

Snedecor, G. W. and Cochran, W. G. (1980). *Statistical Methods*. The Iowa State University Press, Ames, Iowa.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, 1:80-83.