Advances in Box-Jenkins Modeling
1. Model Construction

KEITH WILLIAM HIPEL

Department of Systems Design, University of Waterloo, Waterloo, Ontario, Canada

ANGUS IAN MCLEOD

Department of Statistics, University of Waterloo, Waterloo, Ontario, Canada

WILLIAM C. LENNOX

Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, Canada

Box-Jenkins modeling of time series data can be improved and simplified by adhering to contemporary modeling procedures. This paper gives the theory and techniques of the application of many recent advances that have been made at the identification, estimation, and diagnostic check stages of model development. The inverse autocorrelation function and the inverse partial autocorrelation function are demonstrated to be useful identification tools for both nonseasonal and seasonal models. Parameters can be estimated more efficiently by employing the modified sum of squares technique. At the estimation stage it is also possible to obtain a maximum likelihood estimate for a Box-Cox power transformation. The Akaike information criterion is introduced to formalize mathematically the concept of model parsimony. When checking for model adequacy, knowing the distribution of the residual autocorrelation allows for a sensitive test for residual whiteness. Diagnostic checks are given for verifying the assumption of homoscedasticity of the model residuals. In practice, heteroscedasticity and nonnormality of the residuals can often be removed by a Box-Cox transformation.

INTRODUCTION

In *Box and Jenkins'* [1970] book on time series analysis they describe a family of linear stochastic models that are now commonly referred to as either Box-Jenkins or Arima (autoregressive integrated moving average) models. This work is in fact a culmination of the research of many prominent statisticians, starting with the pioneering work of *Yule* [1927].

When applying a Box-Jenkins model, or in general any type of stochastic model, to a particular problem it is recommended that the three stages of model development be adhered to [*Box and Jenkins,* 1970; *Box and Tiao,* 1973]. The first step is to identify the form of model that may fit the given data. At the estimation stage the model parameters are calculated by employing the method of maximum likelihood. Then the model is checked for possible inadequacies. If the diagnostic checks reveal serious anomalies, appropriate model modifications can be made by repeating the identification and estimation stage.

Since 1970 there have been numerous theoretical and technical application advances in Arima modeling. The purpose of this paper is to show recent innovations that have been made at the identification, estimation, and diagnostic check stages for both seasonal and nonseasonal Arima models. In an accompanying paper labeled part 2 [*McLeod et al.,* 1977] the utility of the methods described in this paper is illustrated by practical applications to actual time series. Although Arima models have previously been applied to hydrologic data [*Carlson et al.,* 1970; *McMichael and Hunter,* 1972; *McKerchar and Delleur,* 1974; *Tao and Delleur,* 1976], the authors maintain that the procedures outlined in this paper substantiate and simplify the modeling process and thereby further enhance the use of Box-Jenkins modeling in water resources. Furthermore, the contemporary modeling methods that are discussed are amenable for use in transfer function-noise model building and intervention analysis [*Hipel et al.,* 1975; *Hipel et al.,* 1977; *Hipel,* 1975; *Box and Tiao,* 1975].

Following a brief theoretical description of Arima difference equations, the three stages of Arima model building are discussed. Theoretical development is given where necessary, and the technique of application of all the methods considered is clearly explained for both nonseasonal and seasonal models. In the appendix the various subdivisions of the three stages of Arima modeling are summarized.

ARIMA PROCESS

Let $z_1, z_2, z_3, \cdots, z_{t-1}, z_t, z_{t+1}, \cdots, z_N$ be a discrete time series measured at equal time intervals. A seasonal Arima model for z_t is written as [*Box and Jenkins,* 1970]

$$\phi(B)\Phi(B^s)\{[(1-B)^d(1-B^s)^D z_t^{(\lambda)}] - \mu\} = \theta(B)\Theta(B^s)a_t \quad (1a)$$

or

$$\phi(B)\Phi(B^s)(w_t - \mu) = \theta(B)\Theta(B^s)a_t \quad (1b)$$

where

- $z_t^{(\lambda)}$ some appropriate transformation of z_t such as a Box-Cox transformation [*McLeod,* 1974; *Box and Cox,* 1964] (no transformation is a possible option);
- t discrete time;
- s seasonal length, equal to 12 for monthly river flows;
- B backward shift operator defined by $Bz_t^{(\lambda)} = z_{t-1}^{(\lambda)}$ and $B^s z_t^{(\lambda)} = z_{t-s}^{(\lambda)}$;
- μ mean level of the process, usually taken as the average of the w_t series (if $D + d > 0$ often $\mu \simeq 0$);
- a_t normally independently distributed white noise residual with mean 0 and variance σ_a^2 (written as NID $(0, \sigma_a^2)$);
- $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ nonseasonal autoregressive (AR) operator or polynomial of order p such

Paper number 6W0879.

that the roots of the characteristic equation $\phi(B) = 0$ lie outside the unit circle for nonseasonal stationarity and the ϕ_i, $i = 1, 2, \cdots, p$ are the nonseasonal AR parameters;

$(1 - B)^d = \nabla^d$ nonseasonal differencing operator of order d to produce nonseasonal stationarity of the dth differences, usually $d = 0, 1$, or 2;

$\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$ seasonal AR operator of order P such that the roots of $\Phi(B^s) = 0$ lie outside the unit circle for seasonal stationarity and the Φ_i, $i = 1, 2, \cdots, P$ are the seasonal AR parameters;

$(1 - B^s)^D = \nabla_s^D$ seasonal differencing operator of order D to produce seasonal stationarity of the Dth differenced data, usually $D = 0, 1$, or 2;

$w_t = \nabla^d \nabla_s^D z_t^{(\lambda)}$ stationary series formed by differencing $z_t^{(\lambda)}$ series ($n = N - d - sD$ is the number of terms in the w_t series);

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \phi_q B^q$ nonseasonal moving average (MA) operator or polynomial of order q such that the roots of $\theta(B) = 0$ lie outside the unit circle for invertibility and θ_i, $i = 1, 2, \cdots, q$ are the nonseasonal MA parameters;

$\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}$ seasonal MA operator of order Q such that the roots of $\Theta(B^s) = 0$ lie outside the unit circle for invertibility and the Θ_i, $i = 1, 2, \cdots, Q$ are the seasonal MA parameters.

The notation $(p, d, q) \times (P, D, Q)_s$ is used to represent the seasonal Arima model of (1). The first set of brackets contains the orders of the nonseasonal operators and the second pair of brackets has the orders of the seasonal operators. For example, a stochastic seasonal noise model of the form $(1, 1, 2) \times (0, 1, 1)_S$ with no data transformation is written as

$$(1 - \phi_1 B)\{[(1 - B)(1 - B^s)z_t] - \mu\} = (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta_1 B^s)a_t \qquad (2)$$

If the model is nonseasonal, only the notation (p, d, q) is needed because the seasonal operators are not present.

When a seasonal model is stationary and requires no differencing (i.e., $D = 0$ and $d = 0$), it is often referred to simply as an Arma (autoregressive moving average) process. The notation $(p, q) \times (P, Q)_s$ is used to represent this type of model. If an Arma model is nonseasonal, the notation (p, q) is used to indicate the orders of the AR and MA operators, respectively. A pure nonseasonal AR process of order p with no differencing is often denoted by AR(p). Likewise, a nonseasonal MA process of order q is sometimes written as MA(q). Of course an AR(p) model can be represented equivalently by the notation $(p, 0)$ or $(p, 0, 0)$, while an MA(q) process can also be denoted by $(0, q)$ or $(0, 0, q)$.

Method of Analysis

1. *Identification*

The purpose of the identification stage is to determine the differencing required to produce stationarity and also the order of both the seasonal and the nonseasonal AR and MA operators for the w_t series. 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.

If at the estimation stage it is decided that the data should be transformed by a Box-Cox transformation, in many cases this does not change the form of the model. 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. Therefore even though it is often not necessary to perform the identification stage for the transformed data, a researcher should be aware that in other 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 transformed data can be properly identified.

In a typical Arima modeling application it is preferable that there be a minimum of about 50 data points in the w_t series in order to get reasonably accurate maximum likelihood estimates (mle) for the parameters. Therefore proceed with the identification stage only if at least the minimum required information is available.

a. Plot of the original series. A visual inspection of a plot of the time series may reveal one or more of the following characteristics: (1) seasonality, (2) trends either in the mean level or in the variance of the series, (3) persistence, (4) long-term cycles, or (5) extreme values and outliers.

b. Autocorrelation function (ACF). The ACF ρ_k measures the amount of linear dependence between observations in a time series that are separated by lag k. *Box and Jenkins* [1970, pp. 32–36] recommend a specific estimation procedure to determine an estimate r_k for ρ_k and also give approximate standard errors for the ACF estimates. To use the ACF in model identification, calculate and then plot r_k against lag k up to a maximum lag of roughly $N/4$.

The first step is to examine a plot of the ACF to detect the presence of nonstationarity in the z_t series. When the data are nonseasonal, failure of the ACF to damp out indicates that nonseasonal differencing is needed. For seasonally correlated 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* [1970, 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.

Once the data have been differenced just enough to produce nonseasonal stationarity for a nonseasonal time series and both seasonal and nonseasonal stationarity for seasonal data, 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 series. The ACF for w_t should not exceed a maximum lag of approximately $n/4$. For a seasonal model a maximum lag of about $5s$ (where $5s < n/4$) is usually sufficient.

If a series is white noise, then r_k is approximately NID $(0, 1/n)$. 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.

When the w_t series is not white noise, then the following general rules may be invoked to help determine the type of model required.

Nonseasonal model: For a pure MA $(0, d, q)$ process, r_k cuts off and is not significantly different from zero after lag q. When the model is pure MA, after lag q [*Bartlett*, 1946],

$$\mathrm{Var}\, r_k \simeq \frac{1}{n}\left(1 + 2\sum_{i=1}^{q} r_i^2\right) \qquad k > q \tag{3}$$

If r_k tails off and does not truncate, this suggests that AR terms are needed to model the time series.

Seasonal model: When the process is a pure MA $(0, d, q) \times (0, D, Q)_s$ model, r_k 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]

$$\mathrm{Var}\, r_k \simeq \frac{1}{n}\left(1 + 2\sum_{i=1}^{q+sQ} r_i^2\right) \qquad k > q + sQ \tag{4}$$

If r_k attentuates 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.

c. Partial autocorrelation function (PACF). The theoretical PACF ϕ_{kk} for an AR process of order k satisfies the Yule-Walker equations [*Box and Jenkins*, 1970, chapter 3]. For model identification calculate and plot the estimates $\hat{\phi}_{kk}$ of ϕ_{kk} against lag k. (The circumflex denotes an estimate of the theoretical statistic below it.) The $\hat{\phi}_{kk}$ are usually calculated for 20 to about 40 lags (where $40 < n/4$). For seasonal models, higher lags of the PACF may be required for identification.

The following general rules may prove helpful for interpreting the PACF of the w_t series.

Nonseasonal model: For a pure AR $(p, d, 0)$ process, $\hat{\phi}_{kk}$ truncates and is not significantly different from zero after lag p. After lag p, $\hat{\phi}_{kk}$ is approximately NID $(0, 1/n)$.

If $\hat{\phi}_{kk}$ tails off, this implies that MA terms are required.

Seasonal model: When the process is a pure AR $(p, d, 0) \times (P, D, 0)_s$ model, $\hat{\phi}_{kk}$ cuts off and is not significantly different from zero after lag $p + sP$. After lag $p + sP$, $\hat{\phi}_{kk}$ is approximately NID $(0, 1/n)$.

If $\hat{\phi}_{kk}$ 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 PACF to truncate at other lags may imply that a nonseasonal MA term is required.

d. Inverse autocorrelation function (IACF). Cleveland [1972] defines the IACF of a time series as the ACF associated with the inverse of the spectral density function of the series. The IACF ρi_k can also be specified in an alternative equivalent manner. Consider the Box-Jenkins $(p, d, q) \times (P, D, Q)_s$ model given by (1*b*). 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

$$\theta(B)\Theta(B^s)(w_t - \mu) = \phi(B)\phi(B^s)a_t \tag{5}$$

To obtain an estimate ri_k for ρi_k at lag k, *Cleveland* [1972] suggests employing either an AR or a smoothed periodogram estimation procedure. If the AR approach is adopted, the first step is to model the w_t series by a finite AR process of order r given by

$$(w_t - \mu) \simeq a_t + \sum_{i=1}^{r} \pi_i(w_{t-i} - \mu) \tag{6}$$

where π_i is the ith AR parameter when the model is written in inverted form. Estimates $\hat{\pi}_i$, where $i = 1, 2, \cdots, r$, for π_i can be determined from the Yule-Walker equations or from the mle of an AR process of order r. The estimates ri_k of the IACF can then be obtained from

$$ri_k = \left(-\pi_k + \sum_{i=1}^{r-k} \hat{\pi}_i\hat{\pi}_{i+k}\right) \Big/ \left(1 + \sum_{i=1}^{r} \hat{\pi}_i^2\right) \tag{7}$$

To utilize the IACF for model identification, calculate and plot ri_k versus lag k. A recommended procedure is to choose about four values of r between 10 and about 40 (where $r < n/4$) and then to select the most representative graph from the set for use in identification. When the model is seasonal the IACF may be calculated for greater than 40 lags if more information is needed for proper identification. Because a selection procedure is required to choose an appropriate IACF plot, the authors suggest that an alternative estimation technique be developed in the future.

If the w_t series is white noise, then ri_k is approximately NID $(0, 1/n)$. When the data are correlated, the following properties of the IACF may be helpful for model identification.

Nonseasonal model: When the process is a pure AR $(p, d, 0)$ model, ri_k cuts off and is not significantly different from zero after lag p. In practice, the authors have found the IACF useful for identifying AR models which have some of the AR parameters equal to zero. At the same lags at which the AR parameters are zero the IACF often has values that are not significantly different from zero. *Cleveland* [1972, pp. 283, 284] substantiates the foregoing fact and shows by a worked example that the PACF fails to detect the AR parameters that are zero in a pure AR process.

If ri_k attenuates, this suggests the presence of a MA component.

Seasonal model: For a pure AR $(p, d, 0) \times (P, D, 0)_s$ process, ri_k truncates and is not significantly different from zero after lag $p + sP$.

If ri_k 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 ri_k to damp out for values between 1 and s, s and $2s$, etc., where decreasing but prominent peaks occur at s, $2s$, $3s$, etc., due to the seasonal MA term.

e. Inverse partial autocorrelation function (IPACF). The IPACF is defined in this paper as the PACF of (5). The 'inverse Yule-Walker equations' are given by

$$\begin{bmatrix} 1 & \rho i_1 & \rho i_2 & \cdots & \rho i_{k-1} \\ \rho i_1 & 1 & \rho i_1 & \cdots & \rho i_{k-2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \rho i_{k-1} & \rho i_{k-2} & \rho i_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi i_{k1} \\ \phi i_{k2} \\ \cdot \\ \cdot \\ \cdot \\ \phi i_{kk} \end{bmatrix} = \begin{bmatrix} \rho i_1 \\ \rho i_2 \\ \cdot \\ \cdot \\ \cdot \\ \rho i_k \end{bmatrix} \tag{8}$$

where ρi_k is the IACF of the w_t series and ϕi_{kj} is the jth coefficient in a MA process of order k such that ϕi_{kk} is the last coefficient.

The coefficient ϕi_{kk} is called the IPACF. To get an estimate $\hat{\phi} i_{kk}$ for ϕi_{kk}, replace ρi_k by the sample IACF ri_k and solve the inverse Yule-Walker equations for $\hat{\phi} i_{kk}$. For model identification plot $\hat{\phi} i_{kk}$ against lag k for the same number of lags as were chosen for the IACF.

Some of the inherent properties of the IPACF are listed below.

Nonseasonal model: If the process is a pure MA $(0, d, q)$ model, $\hat{\phi} i_{kk}$ truncates and is not significantly different from zero after lag q. After lag q, $\hat{\phi} i_{kk}$ is approximately NID $(0, 1/n)$.

When $\hat{\phi} i_{kk}$ dies off rather than cuts off, this suggests that AR terms are required.

Seasonal model: For a pure MA $(0, d, q) \times (0, D, Q)_s$ model, $\hat{\phi}i_{kk}$ truncates and is not significantly different from zero after lag $q + sQ$. After lag $q + sQ$, $\hat{\phi}i_{kk}$ is approximately NID $(0, 1/n)$.

If $\hat{\phi}i_{kk}$ 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.

f. Cumulative periodogram white noise test. As was mentioned previously, the ACF is an accepted means of checking whether the given data are white noise. The PACF, IACF, and IPACF can also be employed in this capacity. However, the cumulative periodogram provides another means of checking for white noise [*Bartlett*, 1966; *Box and Jenkins*, 1970, pp. 294–298].

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

g. Summary of the identification techniques. A plot of the original data portrays an overall view of how the time series is generally behaving. However, the 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.

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 this procedure because it is easier and more accurate to determine the proper Arima model to estimate. It is recommended that all the identification plots be programed for instantaneous display on a cathode ray terminal. In this way the identification stage can usually be completed in 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.

2. *Estimation*

a. Maximum likelihood estimates (mle) for the model parameters. *Box and Jenkins* [1970, chapter 7] suggest that the approximate mle for the Arima model parameters be obtained by employing the unconditional sum of squares method. When using this technique the unconditional sum of squares function is minimized to get least squares parameter estimates.

Recently, *McLeod* [1976*a*] has described a modified sum of squares method which provides parameter estimates that are closer approximations than those of *Box and Jenkins* [1970, chapter 7] to the exact maximum likelihood estimates. The modified sum of squares function is minimized in order to obtain the improved parameter estimates.

By utilizing simulation experiments, *McLeod* [1976*a*] demonstrates inherent assets of the modified estimation procedure. Some general conclusions regarding the advantages of the modified sum of squares method over the unconditional sum of squares technique are that (1) the modified method is more efficient, (2) the modified method gives better parameter estimates for shorter series, and (3) significantly improved parameter estimates are obtained for MA parameters. An additional advantage of the modified sum of squares method is that it is an exact mle procedure for AR processes.

Various optimization techniques are available to minimize functions such as the unconditional sum of squares function and the modified sum of squares function. Some of the optimization algorithms that have been extensively applied include (1) the Gauss linearization [*Draper and Smith*, 1966, chapter 10], (2) the steepest descent [*Draper and Smith*, 1966, chapter 10], (3) the Marquardt algorithm (combination of (1) and (2)) [*Marquardt*, 1963], and (4) conjugate directions [*Powell*, 1964, 1965]. *McLeod* [1976*a*] recommends the use of conjugate directions to minimize the modified sum of squares function. This approach is employed for obtaining parameter estimates for the applications in part 2 [*McLeod et al.*, 1977]. Improvements of the modified procedure over the unconditional sum of squares method are illustrated by the worked examples.

At the estimation stage, estimates are almost always calculated for the AR and MA parameters and σ_a^2 unless the exact value of a parameter is known in advance. For this situation the known parameter can be fixed, and only the remaining parameters are estimated. This is the usual approach taken for the parameter μ of the w_t series. In practice the mle for μ rarely differs from the arithmetic mean of the w_t series, and μ is usually assigned this value. If the data are differenced at least once either seasonally or nonseasonally, then μ usually has a value of zero. However, when it is suspected that a trend component is present, μ can be estimated [*Box and Jenkins*, 1970, pp. 91–93].

b. Box-Cox transformations. In Box-Jenkins modeling the residual a_t are assumed to be independent, homoscedastic (i.e., variance is a constant), and usually normally distributed. The independence assumption is the most important of all, and its violation can cause drastic consequences [*Box and Tiao*, 1973, p. 522]. However, if the constant variance and normality assumptions are not true, they are often reasonably well satisfied when the observations z_t are transformed by a Box-Cox transformation [*McLeod*, 1974, p. 14; *Box and Cox*, 1964].

The normality assumption of the residuals is usually not critical for obtaining good parameter estimates. As long as the a_t are independent and possess finite variance, reasonable estimates (called Gaussian estimates) of the parameters can be obtained [*Hannan*, 1970, p. 372]. *McLeod* [1974, pp. 76–85] demonstrates this fact by Monte Carlo experiments. Simulated data from specified models with known parameters are generated for the a_t distributed as uniform, double exponential, contaminated normal, and normal. For the first three cases, Gaussian estimates of the parameters for the generating model fit to the simulated data are very close to the known values. Of course this is also true for the mle of the parameters for the normal case.

In practice it is advantageous to satisfy the normality assumption reasonably well. First, it can be expected that parameter estimates will be at least slightly improved if a suitable transformation of the z_t is reflected to the a_t by causing them to become approximately normally distributed. Second, without the normality assumption, calculation of confidence intervals for the forecasted data would be impossible for practical use. Finally, if both heteroscedasticity and nonnormality of the residuals are present, then both these flaws can often be rectified simultaneously by a suitable Box-Cox transformation.

Consider power transformations of the form

$$z_t^{(\lambda)} = \lambda^{-1}[(z_t + \text{const})^\lambda - 1] \qquad \lambda \neq 0$$
$$z_t^{(\lambda)} = \ln (z_t + \text{const}) \qquad \lambda = 0 \qquad (9)$$

where const is a constant. The log likelihood of all the AR and MA parameters, σ_a^2, μ (if this parameter is chosen to be estimated), λ, and const, is approximately [*McLeod,* 1974, p. 14]

$$Le \simeq -\frac{n}{2} \ln \frac{\text{mss}}{n} + (\lambda - 1) \sum_{t=d+sD+1}^{N} \ln (z_t + \text{const}) \qquad (10)$$

where mss is the modified sum of squares [*McLeod,* 1976*a*]. When using the estimation procedure of *Box and Jenkins* [1970, chapter 7], the mss is replaced by the sum of squares.

Various approaches are available when determining the values of a Box-Cox transformation. Sometimes it is known in advance that the time series observations of a given phenomenon require a certain type of transformation. For this situation it is appropriate to specify fixed values for both λ and the constant prior to the identification and estimation stages. For example, the authors have found in practice that it is often necessary to transform average monthly river flow data by using natural logarithms. Therefore λ is set equal to zero, and the constant is also assigned a value of zero if no zero observations are present in the series. If there are one or more zero values in the series, the constant can be given a small positive value so that it is possible to take natural logarithms.

In certain instances a chosen standardized transformation, such as a natural logarithm or a square root, may fail to remove heteroscedasticity and/or nonnormality of the residuals. It may therefore be desirable to compute a mle of λ. The first step is to choose a value for the constant that causes all the values of the given series to be greater than zero. Then λ can be estimated simultaneously with the other model parameters.

Calculation of the best Box-Cox transformation for a particular model involves a significant increase in computer time in comparison with the computer time used when λ is not estimated for that model. Therefore it is not recommended that λ be estimated unless diagnostic checks for the residuals indicate that the normality and/or constant variance assumptions are not satisfied.

When it is desired to economize on computer time or when a computer package does not have the capability of estimating λ, it is still possible to select reasonable values for this parameter. Assign the constant a value such that all values of the series have magnitudes greater than zero. Then calculate the log likelihood for, say, $\lambda = 0$, ± 0.5, and ± 1.0, and choose the λ value that gives the largest likelihood.

c. Akaike information criterion (AIC). *Box and Jenkins* [1970] stress the need to use as few model parameters as possible (i.e., the model should be parsimonious) so that the model passes all the diagnostic checks. The AIC [*Akaike,* 1974] is a mathematical formulation of the parsimony criterion of model building.

When there are several competing models to choose from, select the model that gives the minimum of the AIC defined by

$$\text{AIC} = -2 \ln (\text{maximum likelihood}) + 2k \qquad (11)$$

where k is the number of AR and MA parameters to estimate. If μ and/or λ are also estimated, then k is increased by 1 for each extra parameter.

In certain instances the use of the AIC replaces the need for hypothesis testing. Therefore the requirement of subjective judgment for choosing the level of significance in hypothesis testing and the use of statistical tables are explicitly formulated as estimation problems. For example, in part 2 [*McLeod et al.,* 1977] it is questioned whether an AR(3) model is appropriate to model the average annual flows of the Saint Lawrence River at Ogdensburg, New York. Because mle possess a limiting normal distribution [*Pierce,* 1972], by using the estimated standard errors and subjectively choosing a level of significance, hypothesis testing can be done for the model parameters. It is shown that ϕ_2 is not significantly different from zero and therefore should be eliminated from the model. Alternatively, the AIC may be employed for this decision process. For this particular example the AIC also selects an AR model of the order of 3 with ϕ_2 constrained to zero in preference to an AR(3) model.

In the recent past the AIC has been applied to various types of stochastic problems. *Akaike* [1974, 1972*b*] has employed the AIC to select a final Arima model among competing Box-Jenkins models fit to the time series under consideration. The AIC has also been utilized to choose the number of independent variables required in a regression analysis [*Akaike,* 1972*a*], to decide upon the number of factors needed in a factor analysis [*Akaike,* 1972*a*], and to determine the order of a Markov chain process [*Tong,* 1975].

Parzen [1974] introduces a criterion for selecting the order of an AR process to fit to a time series if the series is generated by an AR scheme of finite order. However, Parzen's criterion can be shown to be asymptotically equivalent to the AIC. Because the AIC can also be used for AR processes as well as for many other types of stochastic models, the authors recommend implementation of the AIC in preference to Parzen's approach.

3. *Diagnostic Checks*

One class of diagnostic checks is devised to test model adequacy by overfitting (section 3*a*). This test assumes that the possible types of model inadequacies are known in advance.

However, most diagnostic tests deal with the residual assumptions in order to determine whether the a_t are independent (section 3*b*), homoscedastic (section 3*c*), and normally distributed (section 3*d*). Residual estimates are needed for the tests used in checking the three aforementioned residual assumptions. The estimates for a_t are automatically calculated at the estimation stage along with the mle for the parameters.

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

a. Overfitting. Overfitting involves fitting a more elaborate model than the one estimated to see if including one or more parameters greatly improves the fit. Extra parameters should be estimated for the more complex model only where it is feared that the simpler model may require more parameters. For example, the PACF and the IACF may possess decreasing but significant values at lags 1, 2, and 9. If an Arma (2, 0) model is originally estimated, then a model to check by overfitting the model would be

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_9 B^9)(z_t - \mu) = a_t \quad (12)$$

A mle estimate of ϕ_9 three or four times its standard error would definitely indicate that the more elaborate model should be selected. For this case the AIC and also the remaining tests in this diagnostic check stage would also point out the complex model as the best one to use. *Box and Newbold* [1971] show another interesting application of overfitting. The practitioner must take care to avoid model redundancy which could occur if the AR and MA components were simultaneously enlarged.

Another method of testing model adequacy by overfitting, which was originally suggested by *Whittle* [1952], is to fit a high-order AR model of order r (where $20 < r < 30$). Suppose the original model has $k = p + P + q + Q$ parameters estimated and estimated residual variance $\hat{\sigma}_a^2(k)$. Then it is shown [*McLeod,* 1974, p. 39] that the likelihood ratio statistic is

$$n \ln (\hat{\sigma}_a^2(k)/\hat{\sigma}_a^2(r)) \simeq \chi^2(r - k) \quad (13)$$

where $\hat{\sigma}_a^2(r)$ is the residual variance estimate for an AR process of order r. If the calculated $\chi^2(r - k)$ from (13) is greater than $\chi^2(r - k)$ from the tables at a chosen significance level, then a model with more parameters is needed.

The likelihood ratio test in (13) can also be used to determine if a model containing fewer parameters gives as good a fit as the full model. Specific examples are given in part 2 [*McLeod et al.,* 1977] for the Saint Lawrence River data. The AIC can also be employed to decide whether a model with less parameters is preferable to a model with more parameters.

b. Tests for whiteness of the residuals. To determine whether the residual a_t are white noise, an appropriate 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, new 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

$$r_k(\hat{a}) = \sum_{t=k+1}^{n} \left(\hat{a}_t \hat{a}_{t-k} \Big/ \sum_{t=1}^{n} \hat{a}_t^2 \right) \quad (14)$$

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

$$\mathbf{r}(\hat{a}) = [r_1(\hat{a}), r_2(\hat{a}), \cdots, r_L(\hat{a})]' \quad (15)$$

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,* 1976*b*] that

$$\mathbf{r}(\hat{a}) \simeq N[0, (1/n)\mathbf{U}] \quad (16)$$

where $\mathbf{U} = 1_L - \mathbf{X}'\mathbf{I}^{-1}\mathbf{X}$, 1_L is the identity matrix, $\mathbf{I} \simeq \mathbf{X}'\mathbf{X}$ is the large-sample information matrix, and $\mathbf{X} = [\psi_{i-js}(\Phi), \psi_{i-j}(\phi), \psi_{i-js}(\Theta), \psi_{i-j}(\theta)]$ are the i, j entries in the four partitions of the $\mathbf{X}$ matrix. The dimensions of the matrices $\mathbf{X}$, $\Psi_{i-js}(\Phi)$, $\psi_{i-j}(\phi)$, $\psi_{i-js}(\Theta)$, and $\psi_{i-j}(\theta)$ 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 a general seasonal Box-Jenkins model, transfer function-noise models, and intervention processes.

There are two useful applications of the RACF distribution theorem. A sensitive diagnostic check is first to plot the RACF along with the asymptotic significance intervals for the RACF that are obtained from the diagonal entries of the matrix $(1/n)\mathbf{U}$. If some of the RACF are significantly different from zero, this may mean that the present model is inadequate. The important RACF to examine are the RACF at the first few lags for a nonseasonal model and the RACF at the first couple of lags and also at lags that are multiples of s for a seasonal model. If the present model is insufficient, a proper model can be selected either by changing the model as suggested by *Box and Jenkins* [1970, p. 299] or by repeating the identification, estimation, and diagnostic check stages of model construction.

A second but less sensitive test is to calculate and to perform a significance test for the portmanteau statistic U_L. 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

$$U_L = n \sum_{l=1}^{L} r_l^2(\hat{a}) \simeq \chi^2(L - P - p - Q - q) \quad (17)$$

where L can usually be given a value from 15 to 25 for nonseasonal models and a value of $4s$ for seasonal processes. 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 appropriate changes must be made. Note that *Box and Jenkins* [1970, p. 503] have subtracted one too many degrees of freedom.

An alternative approach that can be used to check for whiteness of the residuals is to examine the cumulative periodogram of the $\hat{a}_t$. However, when considering the residuals it should be remembered that this test is known to be inefficient. Often the cumulative periodogram test fails to indicate model inadequacy due to dependence of the residuals unless the model is a very poor fit to the given data.

c. Homoscedasticity checks of the residuals. The following tests described by *McLeod* [1974] are useful for determining whether a transformation of the data is needed by checking for changes in variance (heteroscedasticity) of the residuals. As was indicated earlier, the variance of the normally independently distributed residuals is assumed to be constant (homoscedastic). Suppose that a_t is NID $[0, \sigma_a^2(t)]$ and that the variance changes with time as $\sigma_a^2(t)$. Let the stochastic random variable ζ_t be NID $(0, \sigma^2)$ and hence have constant variance. Suppose then that

$$a_t = \exp \{(\chi/2)[K(t) - \bar{K}]\} \zeta_t \quad (18)$$

where χ is some constant to be estimated, $K(t)$ is a function of time to be specified, and $\bar{K}$ is the mean of $K(t)$ and equals $n^{-1}\sum_{t=1}^{n} K(t)$. The variance of the a_t residuals is then

$$\sigma_a^2(t) = E\{\exp [\chi(K(t) - \bar{K})] \zeta_t^2\} = \exp \{\chi[K(t) - \bar{K}]\} \sigma^2 \quad (19)$$

It can be shown [*McLeod,* 1974, p. 46] that the natural logarithm of the likelihood Lh for σ^2 and χ is

$$Lh = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{n} \{\exp [-\chi(K(t) - \bar{K})]a_t^2\} \quad (20)$$

and

$$\frac{\partial Lh}{\partial \chi} = \frac{1}{\sigma^2} \sum_{t=1}^{n} \{K(t) \exp [-\chi(K(t) - \bar{K})]a_t^2\} \quad (21)$$

Solve $\partial Lh/\partial\sigma^2 = 0$ exactly for σ^2, and substitute for σ^2 into (21). Equation (21) is set equal to zero, and the residual estimates $\hat{a}_t$ obtained from the estimation stage are used for a_t. This equation is solved for a mle of χ by using the Newton-Raphson method with an initial value of $\chi = 0$.

In order to carry out a test of the hypothesis, the first step is to postulate the null hypothesis that $\chi = 0$ and therefore to assume that the residuals have constant variance. The alternative hypothesis is that the residuals are heteroscedastic and that $\chi \neq 0$. By putting $K(t) = t$ in the previous equations it is possible to test for trends in variance of the residuals over time. If $K(t) = w_t - \hat{a}_t$, then one can check for changes of variance depending on the current level of the series. A likelihood ratio test of the null hypothesis is obtained by computing the mle of χ and comparing it with its standard error. The variance for the mle $\hat{\chi}$ for χ is calculated by using the equation

$$\text{Var } \hat{\chi} = -1/(\partial^2 Lh/\partial\chi^2) \tag{22}$$

Because the mle for χ is asymptotically normally distributed, after a level of significance is chosen it is a straightforward procedure to determine whether to accept or to reject the null hypothesis. This test is also valid for transfer function-noise, intervention, and regression models. In regression models the test for heteroscedasticity can indicate whether an important covariate is missing [*Anscombe,* 1961; *Pierce,* 1971].

If model inadequacy is revealed by either of the tests, a simultaneous estimation procedure can be used to estimate the seasonal and nonseasonal AR and MA parameters σ^2 and χ. This would involve an enormous amount of computer time. However, in practice, the Box-Cox transformation described in the estimation stage will often stabilize the variance.

d. Tests for normality of the residuals. Many standard tests are available to check whether data are normally distributed. For instance, 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 [*Daniel and Wood,* 1971].

Normally distributed data should possess no significant skewness. The skewness g_1 of the residuals is calculated from

$$g_1 = \left(\frac{1}{n}\sum_{t=1}^{n}\hat{a}_t^3\right) \Big/ \left(\frac{1}{n}\sum_{t=1}^{n}\hat{a}_t^2\right)^{3/2} \tag{23}$$

It can be shown that g_1 is approximately N (0, $6/n$).

If the data are normally distributed they should not have a significant g_2 kurtosis coefficient that is given as

$$g_2 = \left(\frac{1}{n}\sum_{t=1}^{n}\hat{a}_t^4\right) \Big/ \left(\frac{1}{n}\sum_{t=1}^{n}\hat{a}_t^2\right)^{2} - 3 \tag{24}$$

The statistic g_2 is approximately N (0, $24/n$). In practice a Box-Cox transformation of the data will often remove any significant skewness or pronounced kurtosis, thereby reinforcing the normality assumption of the residuals.

CONCLUSIONS

Various new procedures are now available to strengthen the three stages of Arima model construction. The IACF and the IPACF are two identification methods that allow for more versatility when designing an Arima model to estimate. All the identification techniques can be used for selecting either a nonseasonal or a seasonal Box-Jenkins model to fit to the data. By employing the modified sum of squares method, more efficient parameter estimates can be obtained. New diagnostic checks are available for checking the independence and homoscedasticity assumptions of the model residuals. If the residuals fail to satisfy the constant variance and/or the normality assumption, often a Box-Cox transformation can rectify the situation. When there are various possible models available for modeling the data, the AIC can be utilized to select the most appropriate model and at the same time to insure model parsimony.

Because of the nature of stochastic problems that occur in water resources the use of Box-Jenkins modeling in this particular field should increase dramatically in the future. The authors recommend that any researcher who deals with Arima modeling use the contemporary procedures that are outlined in this paper. Since all the techniques can be programed, this means that for practical applications the methods can be implemented easily.

Often the expenses incurred when developing an Arima model are insignificant when compared to the costs of various types of applications of the model and also the penalty costs that can arise if an incorrect decision is made due to modeling nature improperly. For example, when a model is employed to simulate data for the economic design of a water resource project, it may be relatively expensive to generate on the computer a sufficient amount of synthetic data. This generated data can be used as input to the economic design of a project such as a reservoir. If the synthetic data are produced by a model that does not fit the data properly, then the final design for the project will be suspect. Furthermore, possible grave economic and social consequences could occur once the project is constructed. Therefore it is advantageous to use the best procedures possible when a model is originally fit to the historical data.

In part 2 [*McLeod et al.,* 1977], Box-Jenkins models are determined for both nonseasonal and seasonal data. The practical applications illustrate the effectiveness of the techniques given in this paper. These methods can also be used for transfer function-noise modeling and intervention analysis.

APPENDIX: STAGES IN BOX-JENKINS MODELING

1. Identification
   - *a.* Plot of the original series
   - *b.* Autocorrelation function (ACF)
   - *c.* Partial autocorrelation function (PACF)
   - *d.* Inverse autocorrelation function (IACF)*
   - *e.* Inverse partial autocorrelation function (IPACF)*
   - *f.* Cumulative periodogram white noise test (to see if the given series is white noise or correlated and if seasonal components are present)
2. Estimation
   - *a.* Maximum likelihood estimates (mle) for the model parameters (modified sum of squares technique)*
   - *b.* Box-Cox transformations*
   - *c.* Akaike information criterion (AIC)*
3. Diagnostic Checks
   - *a.* Overfitting*
   - *b.* Tests for whiteness of the residuals (examination of the residual autocorrelation function (RACF) including the distribution of the RACF estimates)*
   - *c.* Homoscedasticity checks of the residuals (test to check for changes in variance of the residuals over time and also changes depending on the current level of the series to see if a Box-Cox transformation is needed)*

d. Tests for normality of the residuals (skewness and kurtosis checks to determine if a Box-Cox transformation is required)

(Recent advances on the sections followed by asterisks are discussed in this paper.)

NOTATION

a_t white noise time series value at time t.
AR(p) AR process of order p.
B backward shift operator.
const additive constant for a Box-Cox transformation.
d order of the nonseasonal differencing operator.
D order of the seasonal differencing operator.
$E(z_t)$ expected value of z_t.
g_1 estimated residual skewness.
g_2 estimated residual kurtosis.
$\mathbf{I}$ large-sample information matrix.
$K(t)$ specified function of time in the homoscedasticity test for the residuals.
$\bar{K}$ mean of $K(t)$.
ln natural logarithms.
Le log likelihood for Box-Cox transformation.
Lh log likelihood in the homoscedasticity checks.
mle maximum likelihood estimates.
MA(q) MA process of order q.
mss modified sum of squares.
n length of w_t series.
NID normally independently distributed random variable.
N (a, b) normally distributed random variable with mean a and variance b.
N length of z_t time series.
p order of the nonseasonal AR operator.
(p, d) nonseasonal Arma model.
(p, d, q) nonseasonal Arima model.
$(p, d) \times (P, Q)$ seasonal Arma model.
$(p, d, q) \times (P, D, Q)_s$ seasonal Box-Jenkins Arima model.
P order of the seasonal AR operator.
q order of the nonseasonal MA operator.
Q order of the seasonal MA operator.
r_k ACF estimate at lag k.
ri_k estimate for the IACF ρi_k.
$r_k(\hat{a})$ ACF of $\hat{a}_t$ at lag k.
$\mathbf{r}(\hat{a})$ vector of ACF for $\hat{a}_t$ up to lag L.
s seasonal length.
t discrete time.
U_L portmanteau statistic calculated from the estimates of the RACF up to lag L.
$(1/n)\mathbf{U}$ covariance matrix for $\mathbf{r}(\hat{a})$.
Var $\hat{\chi}$ variance of $\hat{\chi}$.
w_t stationary series formed by differencing the $z_t^{(\lambda)}$ series.
$\mathbf{X}$ matrix used in the calculation of the information and covariance matrix.
z_t discrete time series value at time t.
$z_t^{(\lambda)}$ transformation of z_t series.
ζ_t random variable that is NID $(0, \sigma^2)$.
$\theta(B)$ nonseasonal MA operator of order q.
θ_i ith nonseasonal MA parameter.
$\Theta(B^s)$ seasonal MA operator of order Q.
Θ_i ith seasonal MA parameter.
λ exponent for Box-Cox transformation.
μ mean level of the w_t series.
π_i ith AR parameter when the model is written in inverted form.
ρ_k ACF at lag k.
ρi_k IACF at lag k.
σ^2 variance of the random variable ζ_t.
σ_a^2 variance of a_t.
$\sigma_a^2(k)$ residual variance of a process with k parameters.
$\sigma_a^2(r)$ residual variance for an AR process of order r.
$\sigma_a^2(t)$ variance of a_t as a function of time.
$\phi(B)$ nonseasonal AR operator of order p.
ϕ_i ith nonseasonal AR parameter.
ϕ_{kk} PACF (kth coefficient for an AR process of order k).
ϕi_{kj} jth coefficient in a MA process of order k.
ϕi_{kk} IPACF (kth coefficient for a MA process of order k).
$\Phi(B^s)$ seasonal AR operator of order P.
Φ_i ith seasonal AR parameter.
χ constant to be estimated in the homoscedasticity test for the residuals.
$\chi^2(k)$ chi-squared random variable with k degrees of freedom.
$\psi_k(\Phi)$ coefficient of B^k in the Maclaurin series expansion of $[\Phi(B^s)]^{-1}$.
$\psi_k(\phi)$ coefficient of B^k in the expansion of $[\phi(B)]^{-1}$.
$\psi_k(\Theta)$ coefficient of B^k in the expansion of $[\Theta(B^s)]^{-1}$.
$\psi_k(\theta)$ coefficient of B^k in the expansion of $[\theta(B)]^{-1}$.
∇^d nonseasonal differencing operator of order d.
∇_s^D seasonal differencing operator of order D.
$\mathbf{1}_L$ identity matrix.

Acknowledgments. The authors appreciate advice given by T. E. Unny of the Civil Engineering Department of the University of Waterloo. The research was supported by the National Research Council of Canada.

REFERENCES

Akaike, H., Automatic data structure search by the maximum likelihood, paper presented at 5th Hawaii International Conference on System Sciences, Inform. Sci. Program, Dep. of Elec. Eng., Univ. of Hawaii, Honolulu, 1972*a*.

Akaike, H., Use of an information theoretic quantity for statistical model identification, paper presented at 5th Hawaii International Conference on System Sciences, Inform. Sci. Program, Dep. of Elec. Eng., Univ. of Hawaii, Honolulu, 1972*b*.

Akaike, H., A new look at the statistical model identification, *IEEE Trans. Automat. Contr., AC-19*(6), 716–723, 1974.

Anscombe, F. J., Examination of residuals, Paper presented at 4th Berkeley Symposium, Berkeley, Calif., 1961.

Bartlett, M. S., On the theoretical specification of sampling properties of autocorrelated time series, *J. Roy. Statist. Soc., Ser. B, 8*, 27–41, 1946.

Bartlett, M. S., *Stochastic Processes*, Cambridge University Press, New York, 1966.

Box, G. E. P., and D. R. Cox, An analysis of transformations, *J. Roy. Statist. Soc., Ser. B, 26*, 211–252, 1964.

Box, G. E. P., and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day Inc., San Francisco, Calif., 1970.

Box, G. E. P., and P. Newbold, Some comments on a paper of Coen, Gomme and Kendall, *J. Roy. Statist. Soc., Ser. A, 2*, 229–240, 1971.

Box, G. E. P., and D. A. Pierce, Distribution of autocorrelations in autoregressive integrated moving average time series models, *J. Amer. Statist. Ass., 65*, 1509–1526, 1970.

Box, G. E. P., and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, Mass., 1973.

Box, G. E. P., and G. C. Tiao, Intervention analysis with applications to economic and environmental problems, *J. Amer. Statist. Ass., 70*, 70–79, 1975.

Carlson, R. F., A. J. A. MacCormick, and D. G. Watts, Application of linear random models to four annual streamflow series, *Water Resour. Res.*, *6*(4), 1070–1078, 1970.

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

Daniel, C., and F. S. Wood, *Fitting Equations to Data,* Wiley-Interscience, New York, 1971.
Draper, N. R., and H. Smith, *Applied Regression Analysis,* John Wiley, New York, 1966.
Granger, C. W. J., and P. Newbold, Forecasting transformed series, *J. Roy. Statist. Soc., Ser. B, 38,* 189–203, 1976.
Hannan, E. J., *Multiple Time Series,* John Wiley, New York, 1970.
Hipel, K. W., Contemporary Box-Jenkins modelling in hydrology, Ph.D. thesis, Univ. of Waterloo, Waterloo, Ont., Canada, 1975.
Hipel, K. W., W. C. Lennox, T. E. Unny, and A. I. McLeod, Intervention analysis in water resources, *Water Resour. Res., 11*(6), 855–861, 1975.
Hipel, K. W., A. I. McLeod, and E. A. McBean, Stochastic modeling of the effects of reservoir operation, *J. Hydrol., 32,* 97–113, 1977.
Marquardt, D. W., An algorithm for least squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math., 11*(2), 431–441, 1963.
McKerchar, A. I., and J. W. Delleur, Application of seasonal parametric linear stochastic models to monthly flow data, *Water Resour. Res., 10*(2), 246–255, 1974.
McLeod, A. I., Contributions to applied time series, masters thesis, Univ. of Waterloo, Waterloo, Ont., Canada, 1974.
McLeod, A. I., Improved Box-Jenkins estimators, report, Dep. of Statist., Univ. of Waterloo, Waterloo, Ont., Canada, 1976*a*.
McLeod, A. I., Distribution of the residual autocorrelations in seasonal Arima models, report, Dep. of Statist., Univ. of Waterloo, Waterloo, Ont., Canada, 1976*b*.
McLeod, A. I., K. W. Hipel, and W. C. Lennox, Advances in Box-Jenkins modeling, 2, Applications, *Water Resour. Res., 13,* this issue, 1977.
McMichael, F. C., and J. S. Hunter, Stochastic modeling of temperature and flow in rivers, *Water Resour. Res., 8*(1), 87–98, 1972.
Parzen, E., Some recent advances in time series modeling, *IEEE Trans. Automat. Contr., AC-19*(6), 723–730, 1974.
Pierce, D. A., Distribution of residual autocorrelation in the regression model with autoregressive moving average errors, *J. Roy. Statist. Soc., Ser. B, 33,* 140–146, 1971.
Pierce, D. A., Least squares estimation in dynamic-disturbance time series models, *Biometrika, 59*(1), 73–78, 1972.
Powell, M. J. D., An efficient method for finding the minimum of a function of several variables without calculating derivatives, *Comput. J., 7,* 155–162, 1964.
Powell, M. J. D., A method for minimizing a sum of squares of nonlinear functions without calculating derivatives, *Comput. J., 8,* 303–307, 1965.
Tao, P. C., and J. W. Delleur, Seasonal and nonseasonal Arma models in hydrology, *J. Hydraul. Div. Amer. Soc. Civil Eng., HY10,* 1541–1559, 1976.
Tong, H., Determination of the order of a Markov chain by Akaike's information criterion, *J. Appl. Probab., 12,* 488–497, 1975.
Whittle, P., Tests of fit in time series, *Biometrika, 39,* 309–318, 1952.
Yule, G. U., On a method of investigating periodicities in disturbed series, with special reference to Wolfer sunspot numbers, *Phil. Trans. Roy. Soc. London, Ser. A, 226,* 267–298, 1927.

(Received June 19, 1975;
revised November 11, 1976;
accepted November 15, 1976.)