Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models

RUEY S. TSAY and GEORGE C. TIAO*

A unified approach for the tentative specification of the order of mixed stationary and nonstationary ARMA models is proposed. For the ARMA models, an iterative regression procedure is given to produce consistent estimates of the autoregressive parameters. An extended sample autocorrelation function based on these consistent estimates is then defined and used for order determination. One of the advantages of this new approach is that it eliminates the need to determine, usually rather arbitrarily, the order of differencing to produce stationarity in modeling time series. Comparisons with other existing identification methods are discussed, and several samples are given.

KEY WORDS: AIC; ARMA models; Extended sample autocorrelations; S-array; Time series; Yule-Walker equations.

1. INTRODUCTION

This article proposes (a) consistent least squares estimates of the autoregressive parameters and (b) a procedure for tentative model specification with respect to a class of nonstationary and stationary autoregressive moving average models.

Consider the ARMA(p, q) model for a univariate time series $\{Z_t\}$,

$$\Phi(B)Z_t = C + \theta(B)a_t, \tag{1.1}$$

where $\Phi(B) = U(B)\phi(B) = 1 - \Phi_1 B - \cdots - \Phi_p B^p$, $U(B) = 1 - U_1 B - \cdots - U_d B^d$, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_{p-d} B^{p-d}$, and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ are polynomials in B; B is the backshift operator such that $BZ_t = Z_{t-1}$; C is a constant; and $\{a_t\}$ is a Gaussian white noise process with mean zero and variance σ_a^2. We shall require that all the zeros of $U(B)$ are on, those of $\phi(B)$ are outside, and those of $\theta(B)$ are on or outside the unit circle, and also that $\Phi(B)$ and $\theta(B)$ have no common factors. Further, we shall assume that Z_t starts at a finite time point t_0 if it is nonstationary. Note that when $U(B) = (1 - B)^d$, (1.1) reduces to the usual ARIMA($p - d$, d, q) model.

In the literature, autoregressive fitting by the ordinary least squares (OLS) method is perhaps the most frequently used approach in time series estimation; see for instance, Box and Jenkins (1976), Durbin (1960), and Whittle (1963). However, as is shown in Tiao and Tsay (1983), for ARMA models the OLS estimates of the autoregressive (AR) parameters in autoregressive fittings are not always consistent. In this article we first propose an iterative regression procedure that yields consistent least squares (LS) estimates for the AR parameters. If the underlying process is stationary, that is, $U(B) = 1$ in (1.1), these new AR estimates are shown to be asymptotically equivalent to the solutions of the estimated generalized Yule-Walker equations. This equivalence, however, breaks down in the nonstationary situation.

We next consider the problem of tentative model specification or identification. If Z_t in (1.1) is stationary, several approaches have been proposed for practical uses: for examples, the sample autocorrelation functions (SACF) and partial autocorrelation functions (SPACF) by Box and Jenkins (1976); the information criterion (AIC) by Akaike (1974); the R-array and S-array by Gray, Kelley, and McIntire (1978); and the corner method recently by Beguin, Gourieroux, and Monfort (1980). The Box-Jenkins approach is useful in specifying purely autoregressive ($q = 0$) or purely moving average ($p = 0$) models; but it is usually difficult to determine the values of p and q for the mixed models. The AIC approach is inconsistent, and the practical implications of this inconsistency have been widely discussed in the literature; see Kashyap (1980), Schwarz (1978), Shibata (1976,1980), and Taniguchi (1980). The R-and-S-array approach is rather complex for practical use, and statistical properties of the statistics involved are largely unknown. The latter problem also exists for the corner method. If Z_t is nonstationary, differencing is usually employed to transform it into a stationary one and the differenced process is then analyzed. However, ambiguity remains in determining the appropriate order of differencing to produce stationarity in many practical applications. Therefore, the sec-

* Ruey S. Tsay is Assistant Professor, Department of Statistics, Carnegie-Mellon University, Pittsburgh, PA 15213. George C. Tiao is W. Allen Wallis Professor of Statistics, Graduate School of Business, University of Chicago, Chicago, IL 60637. The authors would like to thank the Associate Editor and two referees for helpful comments. This research is partially supported by a grant from the Wisconsin Alumni Research Foundation.

March 1984, Volume 79, Number 385
Theory and Methods Section

ond and main purpose of this article is to propose a unified procedure to resolve the preceding identification problems for both nonstationary and stationary ARMA models. Based on the consistent AR estimates produced by iterated regressions, an extended sample autocorrelation function (ESACF) is defined and then used for tentative model specification. Statistical properties of the ESACF are discussed. One of the advantages of the proposed ESACF approach is that it eliminates the need to determine, usually rather arbitrarily, the order of differencing to achieve stationarity.

In Section 2, the iterative AR regression procedure is introduced, and the main properties of the corresponding iterated AR estimates are stated without proofs. Based on the consistency property of these estimates, Section 3 defines the ESACF and shows the "cutting-off" property when the correct AR order is fitted. It also describes the asymptotic behavior of the ESACF when the AR order is overfitted. The model identification procedure (henceforth referred to as the ESACF approach) is given in Section 4 with several examples. Section 5 discusses in detail the properties of the iterated AR estimates, and Section 6 derives the properties of ESACF. Finally, some further discussion of the ESACF approach and comparison with other existing identification procedures are presented in Section 7.

For simplicity, we assume throughout that the constant C in (1.1) is zero and employ the following notations: $e_t = \theta(B)a_t$ is the pure MA part of (1.1), $\hat{\beta}$ is the LS estimate of β, $\mathbf{A}'$ and $\det(\mathbf{A})$ are, respectively, the transpose and determinant of the matrix $\mathbf{A}$, $\doteq$ denotes asymptotic equivalence in probability, and $\xrightarrow{P}$ means convergence in probability.

2. ITERATED REGRESSIONS

Suppose n observations are available from the ARMA(p, q) process

$$Z_t = \sum_{l=1}^{p} \Phi_l Z_{t-l} - \sum_{j=1}^{q} \theta_j a_{t-j} + a_t. \qquad (2.1)$$

Our goal is to obtain consistent LS estimates of the AR parameters Φ_l's. First, suppose an ordinary AR(p) regression is fitted to the data. This regression will be written as

$$Z_t = \sum_{l=1}^{p} \Phi_{l(p)}{}^{(0)} Z_{t-l} + e_{p,t}{}^{(0)}, \quad t = p+1, \ldots, n, \qquad (2.2)$$

where the superscript (0) means the ordinary autoregression, the subscript (p) denotes the order of the AR fitting, and $e_{p,t}{}^{(0)}$ is the corresponding error term. It is known that the OLS estimates $\hat{\Phi}_{l(p)}{}^{(0)}$ are consistent for Φ_l, that is,

$$\hat{\Phi}_{l(p)}{}^{(0)} \xrightarrow{P} \Phi_l, \quad l = 1, \ldots, p$$

if the underlying process follows either an AR(p) model ($q = 0$) or a purely nonstationary ARMA(p, q) model ($\phi(B) = 1$). This result was given by Mann and Wald (1943) for the stationary case, while that for the nonstationary situation can be found in Tiao and Tsay (1983). In the latter article, it was also shown that $\hat{\Phi}_{l(p)}{}^{(0)}$ will be inconsistent for other models.

Next, consider the case that $\hat{\Phi}_{l(p)}{}^{(0)}$ are inconsistent. Since the estimated residuals $\hat{e}_{p,t}{}^{(0)}$ of (2.2), where

$$\hat{e}_{p,t}{}^{(0)} = Z_t - \sum_{l=1}^{p} \hat{\Phi}_{l(p)}{}^{(0)} Z_{t-l},$$

are not white noise even for large n, the lagged values $\hat{e}_{p,t-j}{}^{(0)}$, $j > 0$, will contain some information about the process Z_t. This has motivated us to define the first iterated AR(p) regression as

$$Z_t = \sum_{l=1}^{P} \Phi_{l(p)}{}^{(1)} Z_{t-l} + \beta_{1(p)}{}^{(1)} \hat{e}_{p,t-1}{}^{(0)} + e_{p,t}{}^{(1)}, \quad t = p+2, \ldots, n, \qquad (2.3)$$

where the superscript (1) means the first iterated regression and $e_{p,t}{}^{(1)}$ denotes the corresponding error term. It is shown later that the LS estimates $\hat{\Phi}_{l(p)}{}^{(1)}$ of this regression are consistent, that is,

$$\hat{\Phi}_{l(p)}{}^{(1)} \xrightarrow{P} \Phi_l, \quad l = 1, \ldots, p$$

if $q \leq 1$ or $\phi(B) = 1$.

Similarly, if $\hat{\Phi}_{l(p)}{}^{(1)}$ are inconsistent for the true AR coefficients, the lagged values of the estimated residuals $\hat{e}_{p,t}{}^{(1)}$ of (2.3), where

$$\hat{e}_{p,t}{}^{(1)} = Z_t - \sum_{l=1}^{P} \hat{\Phi}_{l(p)}{}^{(1)} Z_{t-l} - \hat{\beta}_{1(p)}{}^{(1)} \hat{e}_{p,t-1}{}^{(0)},$$

also retains some useful information about Z_t. We thus define the second iterated AR(p) regression as

$$Z_t = \sum_{l=1}^{P} \Phi_{l(p)}{}^{(2)} Z_{t-l} + \beta_{1(p)}{}^{(2)} \hat{e}_{p,t-1}{}^{(1)} + \beta_{2(p)}{}^{(2)} \hat{e}_{p,t-2}{}^{(0)} + e_{p,t}{}^{(2)}, \quad t = p+3, \ldots, n, \qquad (2.4)$$

where $\hat{e}_{p,t-2}{}^{(0)}$ and $\hat{e}_{p,t-1}{}^{(1)}$ are, respectively, the estimated residuals of (2.2) and (2.3); and it will be shown that

$$\hat{\Phi}_{l(p)}{}^{(2)} \xrightarrow{P} \Phi_l, \quad l = 1, \ldots, p$$

if $q \leq 2$ or $\phi(B) = 1$, and so on.

In practice, the true orders (p, q) of an ARMA(p, q) process are usually unknown, particularly at the identification stage of the analysis. The preceding considerations have led us to consider a general set of iterated regressions. Specifically, the jth iterated AR(k) regression of a time series $\{Z_t\}$ is defined as

$$Z_t = \sum_{l=1}^{k} \Phi_{l(k)}{}^{(j)} Z_{t-l} + \sum_{i=1}^{j} \beta_{i(k)}{}^{(j)} \hat{e}_{k,t-i}{}^{(j-i)} + e_{k,t}{}^{(j)},$$

$$t = k + j + 1, \ldots, n; \quad j = 0, \ldots; \quad k = 1, 2, \ldots, \qquad (2.5)$$

where

$$\hat{e}_{k,t}^{(i)} = Z_t - \sum_{l=1}^{k} \hat{\Phi}_{l(k)}^{(i)} Z_{t-l} - \sum_{h=1}^{i} \hat{\beta}_{h(k)}^{(i)} \hat{e}_{k,t-h}^{(i-h)} \tag{2.6}$$

is the estimated residual of the ith iterated AR(k) regression and $\hat{\Phi}_{l(k)}^{(i)}$'s and $\hat{\beta}_{h(k)}^{(i)}$'s are the corresponding LS estimates.

For convenience to the reader, we state here the main properties of the iterated AR estimates $\hat{\Phi}_{l(k)}^{(j)}$ and postpone all the proofs to Section 5. First, the iterated AR estimates satisfy the recursion

$$\hat{\Phi}_{l(k)}^{(j)} = \hat{\Phi}_{l(k+1)}^{(j-1)} - \hat{\Phi}_{l-1(k)}^{(j-1)} \hat{\Phi}_{k+1(k+1)}^{(j-1)} / \hat{\Phi}_{k(k)}^{(j-1)} \tag{2.7}$$

where $\hat{\Phi}_{0(k)}^{(j-1)} = -1$, $l = 1, \ldots, k$; $k \geq 1$ and $j \geq 1$. The implication of (2.7) is that the AR estimates of the jth iterated AR(k) regression in practice can be recursively computed from the OLS estimates of stepwise AR(k), AR($k+1$), . . . , AR($k+j$) fittings.

Next, the consistency properties of the iterated AR estimates can be summarized in the following two theorems.

Theorem 2.1. Suppose that Z_t follows a nonstationary ARMA(p, q) model in (1.1). Then

$$\hat{\Phi}_{l(d)}^{(j)} = U_l + O_p(n^{-1}), \quad l = 1, \ldots, d; \quad j \geq 0.$$

The practical implication of Theorem 2.1 is important. It implies that for a nonstationary ARMA(p, q) process $\{Z_t\}$ in (1.1), the LS estimates $\hat{\Phi}_{l(d)}^{(j)}$ of any jth iterated AR(d) regression are consistent for the nonstationary AR coefficients U_l. This provides a useful procedure to specify the nonstationary factor in practical time series modeling. Examples of this respect are given in Section 4.

Theorem 2.2. Suppose that Z_t is an ARMA(p, q) process, stationary or not. Then

$$\hat{\Phi}_{l(k)}^{(j)} = \Phi_l + O_p(n^{-1/2}), \quad l = 1, \ldots, k$$

if $k \geq p$ and $j = q$, or $k = p$ and $j > q$, where it is understood that $\Phi_l = 0$ for $l > p$. Also, $O_p(n^{-1/2})$ becomes $O_p(n^{-1})$ if $\phi(B) = 1$ and $k = p$.

3. THE EXTENDED SAMPLE AUTOCORRELATION FUNCTIONS

We now turn to the problem of tentative order specification in time series modeling. Theorem 2.2 shows that consistent estimates of the AR parameters in an ARMA(p, q) process can be obtained by iterated regressions. Using these consistent estimates, we now define the extended sample autocorrelation function (ESACF) $r_{j(k)}$ in the following manner:

1. If $p = 0$, then Z_t follows an MA(q) model and its SACF cuts off asymptotically at lag q. That is,

$$r_{j(0)} \doteq 0 \quad \text{if} \quad j > q \quad \text{and} \quad p = 0, \tag{3.1}$$

where the subscript (0) denotes the ordinary SACF of Z_t and j is the lag.

2. If $p = 1$, then Z_t is an ARMA(1, q) process. In this case, if a consistent estimate $\tilde{\Phi}_1$ of Φ_1 can be found, then $W_t = Z_t - \tilde{\Phi}_1 Z_{t-1}$ will asymptotically follow a pure MA(q) model, and its SACF will have the property in (3.1). However, in obtaining a consistent estimate of Φ_1, the number of iterations needed in the iterative regression procedure depends on the true value of q, which in practice is unknown. We are thus led to the following considerations:

(i) If $q = 0$, then, by Theorem 2.2, $\hat{\Phi}_{1(1)}^{(j)} \xrightarrow{P} \Phi_1$ for all $j \geq 0$ and in particular the transformed series $\{W_{1,t}^{(0)}\}$, where $W_{1,t}^{(0)} = Z_t - \hat{\Phi}_{1(1)}^{(0)} Z_{t-1}$, is asymptotically white noise so that $r_s(W_{1,t}^{(0)}) \doteq 0$ for $s > 0$ where $r_s(X_t)$ denotes the SACF of X_t at lag s.

(ii) If $q = 1$, then $\hat{\Phi}_{1(1)}^{(j)} \xrightarrow{P} \Phi_1$ for $j \geq 1$ and the transformed series $\{W_{1,t}^{(1)}\}$, where $W_{1,t}^{(1)} = Z_t - \hat{\Phi}_{1(1)}^{(1)} Z_{t-1}$, follows the MA(1) model $W_{1,t}^{(1)} = (1 - \theta_1 B) a_t$ for sufficiently large n. This implies that $r_s(W_{1,t}^{(1)}) \doteq 0$ for $s > 1$.

(iii) More generally, $\hat{\Phi}_{1(1)}^{(j)} \xrightarrow{P} \Phi_1$ for $j \geq q$ and the transformed series $\{W_{1,t}^{(q)}\}$, where $W_{1,t}^{(q)} = Z_t - \hat{\Phi}_{1(1)}^{(q)} Z_{t-1}$, follows an MA($q$) model so that $r_s(W_{1,t}^{(q)}) \doteq 0$ for $s > q$.

We can now define the first ESACF of Z_t as

$$r_{j(1)} = r_j(W_{1,t}^{(j)}), \tag{3.2}$$

where the subscript (1) denotes the first extended SACF. One can then easily see that

$$\begin{aligned} r_{j(1)} &\doteq 0 \quad \text{for} \quad j > q \quad \text{and} \quad p = 1 \\ &\neq 0 \quad \text{for} \quad j = q \quad \text{and} \quad p = 1. \end{aligned} \tag{3.3}$$

3. In general, for any nonnegative integer k we define the kth ESACF of Z_t as

$$r_{j(k)} = r_j(W_{k,t}^{(j)}) \tag{3.4}$$

where $W_{k,t}^{(j)} = Z_t - \sum_{l=1}^{k} \hat{\Phi}_{l(k)}^{(j)} Z_{t-l}$. Then, if the true model is an ARMA(p, q) we have that asymptotically $W_{p,t}^{(j)}$ follows an MA(q) model for $j \geq q$ so that

$$\begin{aligned} r_{j(k)} &\doteq 0 \quad \text{for} \quad j > q \quad \text{and} \quad k = p \\ &\neq 0 \quad \text{for} \quad j = q \quad \text{and} \quad k = p. \end{aligned} \tag{3.5}$$

A formal proof of (3.5) is given in Lemma 6.1 of Section 6.

Comparing (3.1) and (3.5), we see that the pth ESACF of an ARMA(p, q) process has exactly the same "cutting-off" property as that of the ordinary SACF of a pure MA process. Moreover, since each individual element of the ESACF is a sample autocorrelation of a transformed series, its sampling properties can be obtained from the known asymptotic results of the SACF. However, except for $k = 0$ the ESACF is *not* the SACF of *any one* transformed series of Z_t. The effect of this feature on model specification is discussed later.

Overfitting

For ARMA(p, q) models with known p, the "cutting-off" feature of the ESACF in (3.5) makes it a useful tool for the tentative specification of the value of q. This feature depends critically on the consistency properties of the iterated AR estimates $\hat{\Phi}_{l(k)}^{(j)}$ established in Theorem 2.2. In practice, the values of p and q are unknown, and complications arise when both the fitted AR order k is greater than the true AR order p and the number of iterations j is greater than the unknown true MA order q. This is referred to as the overfitting problem. In Section 6, we prove that the effect of overfitting $k - p > 0$ autoregressive lags will in general lead to an increase in the order of the moving average polynomial of the transformed series $\{W_{k,t}^{(j)}\}$, and the number of additional terms is given by $\min\{k - p, j - q\}$. Specifically, for the ESACF, we show that for an ARMA(p, q) model

$$r_{j(k)} \doteq \begin{cases} c(k-p, j-q), & 0 \le j-q \le k-p \\ 0, & j-q > k-p \le 0, \end{cases} \quad (3.6)$$

where $c(k - p, j - q)$ is some nonzero constant or a continuous random variable bounded between -1 and 1.

4. TENTATIVE MODEL IDENTIFICATION

The asymptotic property of the ESACF $r_{j(k)}$ given by (3.6) can now be exploited to help tentatively identify ARMA(p, q) models in practice. For this purpose it is useful to arrange the $r_{j(k)}$'s in a two-way table as shown in Table 1, in which the first row gives the SACF, the second row gives the 1st ESACF, and so on. The rows are numbered 0, 1, 2, . . . , to signify the AR order and the columns in a similar way for the MA order. To illustrate the use of the table, suppose the true model is an ARMA(1, 2). For the SACF, it is well known that asymptotically $r_{j(0)} \neq 0$ for $j \ge 2$. Now from (3.6) with $p = 1$ and $q = 2$, we see that (i) when $k = 1$, $r_{j(1)} \doteq 0$ for $j \ge 3$, (ii) when $k = 2$, $r_{j(2)} \doteq 0$ for $j \ge 4$, and so on. The full situation is shown in Table 2, where X denotes a nonzero value, 0 is zero, and * means a value between -1 and 1. The zero values are seen to form a triangle with boundaries given by the two lines $k = 1$ and $j - k = 2$. The row and column coordinates of the vertex correspond precisely to the AR and MA order, respectively.

In general, we are thus led to search from the ESACF

Table 1. The ESACF Table

	MA					
AR	0	1	2	3	·	·
0	$r_{1(0)}$	$r_{2(0)}$	$r_{3(0)}$	$r_{4(0)}$	·	·
1	$r_{1(1)}$	$r_{2(1)}$	$r_{3(1)}$	$r_{4(1)}$	·	·
2	$r_{1(2)}$	$r_{2(2)}$	$r_{3(2)}$	$r_{4(2)}$	·	·
3	$r_{1(3)}$	$r_{2(3)}$	$r_{3(3)}$	$r_{4(3)}$	·	·
·	·	·	·	·	·	·
·	·	·	·	·	·	·

*Table 2. The Asymptotic ESACF Table for an ARMA(1, 2) Model Where X Denotes a Nonzero Value and * 's Denote No Pattern Terms*

	MA							
AR	0	1	2	3	4	5	6	7
0	*	X	X	X	X	X	X	X
1	*	X	0	0	0	0	0	0
2	*	X	X	0	0	0	0	0
3	*	X	X	X	0	0	0	0
4	*	X	X	X	X	0	0	0

table the vertex of a triangle of asymptotic "zero" values having boundary lines $k = c_1 \ge 0$ and $j - k = c_2 \ge 0$, and tentatively identify $p = c_1$ and $q = c_2$ as the order of the ARMA model. In practice, for finite samples, the $r_{j(k)}$'s will not be zero. The asymptotic variance of the $r_{j(k)}$'s can be approximately obtained by using Bartlett's formula. As a crude but simple approximation, we may use the value $(n - k - j)^{-1}$ on the hypothesis that the transformed series $W_{k,t}^{(j)}$ of (3.4) is white noise to estimate the variance of $r_{j(k)}$. Of course, it is understood that this simple approximation might underestimate the variance of $r_{j(k)}$ and a further study on this subject is needed in the future. As a preliminary but informative guide for model specification, the ESACF table may be supplemented by an analogous table consisting of indicator symbols X denoting values greater or less than ± 2 standard deviations and 0 for in-between values. We illustrate this approach with several real and simulated examples.

Example 1. Consider the Series C in Box and Jenkins (1976) consisting of 226 temperature readings taken every minute from a chemical process. Table 3(a) gives the ESACF and Table 3(b) the corresponding indicator symbols for this data set. The "triangular cutting-off" behavior starting at the third row clearly suggests an AR(2) model.

Table 3. The ESACF of Series C

	MA								
AR	0	1	2	3	4	5	6	7	8
					a. The ESACF Table				
0	.98	.94	.90	.85	.80	.75	.69	.64	.58
1	.81	.66	.55	.48	.43	.38	.34	.28	.25
2	−.04	−.03	−.12	−.06	.02	−.01	.07	−.04	−.12
3	−.50	.01	−.07	−.11	−.01	.00	.03	−.03	−.10
4	−.25	−.27	−.05	−.11	−.01	.03	.00	−.02	−.09
5	−.48	.28	−.29	−.07	.04	−.05	−.00	−.01	−.08
					b. The Indicator Symbols				
0	X	X	X	X	X	X	X	X	X
1	X	X	X	X	X	X	X	X	X
2	0	0	0	0	0	0	0	0	0
3	X	0	0	0	0	0	0	0	0
4	X	X	0	0	0	0	0	0	0
5	X	X	X	0	0	0	0	0	0

For this example, a brief discussion is in order. After an AR(2) model is specified, one can then easily obtain the estimates of the autoregressive parameters from the AR(2) regression. Specifically, for this example $\hat{\Phi}_{1(2)}^{(0)} = 1.81$ and $\hat{\Phi}_{2(2)}^{(0)} = -.82$ so that $\hat{\Phi}_2^{(0)}(B) \simeq (1 - B)(1 - .8B)$, strongly indicating that $\Phi(B)$ contains the factor $(1 - B)$. Note that the ESACF approach does *not* require any differencing of Z_t before specifying an AR(2) model containing the factor $(1 - B)$. On the other hand, the S-array approach by Gray, Kelly, and McIntire (1978) might lead the user to specify an ARMA(2, 1) model for Series C if it is applied directly to the original observations. In fact, these authors used this data set to emphasize the necessity of stationarity requirement in the S-array approach and to illustrate the need of special attention in interpreting the S-array. Of course, for this example, the AR(2) model would also be identified from the sample PACF if they are computed by stepwise AR fittings, but the PACF approach only works for pure AR models.

Example 2. We next consider a nonstationary ARMA(4, 1) model where $\Phi(B)$ contains two unit roots and a pair of complex roots also lying on the unit circle. One hundred observations were generated from the process

$$(1 - B)^2(1 - \sqrt{2}B + B^2)Z_t = (1 - .5B)a_t, \quad \sigma_a^2 = 1.0, \quad (4.1)$$

and the ESACF are given in Table 4. The apparent triangular pattern starting at the fifth row of the table does indeed suggest an ARMA(4, 1) model.

For this example, it is of interest to examine the coefficients of the iterated AR(4) fittings after an ARMA(4, 1) model is specified. These estimated coefficients $\hat{\Phi}_{l(k)}^{(j)}$ for $k = 4$; $l = 1, \ldots, 4$; and $j = 0, \ldots, 4$ are given in Table 5. The consistency property of the AR estimates in the iterated regression for $j \geq 1$ is clearly shown.

For this nonstationary example, the S-array approach and the more recently proposed generalized partial autocorrelation functions approach (Woodward and Gray 1981) would require the following steps to specify the model: (a) compute the S-array of the original observations to pick up a $(1 - B)$ factor; (b) compute the S-array of $(1 - B)Z_t$ to specify the $(1 - B)(1 - \sqrt{2}B + B^2)$ factor; (c) compute the S-array of $(1 - B)^2(1 - \sqrt{2}B + B^2)Z_t$ to identify the MA order. Thus it is clear from this and the previous examples that, for nonstationary data, the S-array approach may require several steps before one arrives at a final model. On the other hand, the approach proposed in this article can handle nonstationary as well as stationary models in a much more direct manner.

Table 4. The ESACF of Example 2

	MA							
AR	0	1	2	3	4	5	6	7
0	.96	.93	.89	.86	.83	.80	.76	.73
1	.72	.09	−.53	−.78	−.55	.01	.54	.75
2	.72	.09	−.58	−.87	−.53	.04	.59	.81
3	.67	.23	−.27	−.44	−.24	.15	.48	.54
4	−.38	−.02	.09	.04	−.11	.06	.07	−.15
5	−.47	−.20	.08	.04	−.11	.05	.05	−.09
6	.08	.31	−.17	.04	.01	−.03	.01	−.05

Table 5. The Estimated Coefficients of the Iterated AR(4) Regressions of Example 2

j	$\Phi_{1(4)}^{(j)}$	$\Phi_{2(4)}^{(j)}$	$\Phi_{3(4)}^{(j)}$	$\Phi_{4(4)}^{(j)}$
0	3.34	−4.67	3.26	−.94
1	3.40	−4.80	3.40	−.99
2	3.40	−4.81	3.40	−.99
3	3.40	−4.81	3.40	−.99
4	3.40	−4.80	3.39	−.99
True values	3.41	−4.83	3.41	−1.00

Example 3. Consider the Series A in Box and Jenkins (1976) consisting of 197 concentration readings taken every two hours from a chemical process. A fitted ARMA(1, 1) model,

$$(1 - .92B)Z_t = 1.45 + (1 - .58B)a_t, \quad \hat{\sigma}_a^2 = .097 \quad (4.2)$$

was obtained by these authors. Table 6 shows the ESACF and the corresponding indicator symbols for the data set. An apparent triangular cutting-off pattern with its vertex at the ARMA(1, 1) position emerges from the indicator symbols except for the three points in the seventh column. However, the numerical values of these three points are only marginally larger than the two standard deviation values that are approximately equal to .15. Thus an ARMA(1, 1) model is also suggested by the ESACF approach. Furthermore, for this model we find from the first

Table 6. The ESACF of Series A

	MA								
AR	0	1	2	3	4	5	6	7	8
					a. The ESACF Table				
0	.57	.50	.40	.36	.33	.35	.39	.32	.30
1	−.39	.04	−.06	−.01	−.07	−.01	.16	−.07	.04
2	−.29	−.27	−.04	.01	−.05	−.01	.17	.03	.04
3	−.50	−.01	.09	−.01	−.01	−.03	.16	−.03	.11
4	−.48	−.02	.08	−.02	−.01	−.04	.14	.03	.09
5	−.39	−.41	−.17	.01	−.17	−.02	.10	−.01	.06
6	−.49	.15	−.18	−.00	−.26	−.06	.09	−.10	.05
7	.19	−.01	.04	.34	.26	−.08	−.23	.03	.01
					b. The Indicator Symbols				
0	X	X	X	X	X	X	X	X	X
1	X	0	0	0	0	0	X	0	0
2	X	X	0	0	0	0	X	0	0
3	X	0	0	0	0	0	X	0	0
4	X	0	0	0	0	0	0	0	0
5	X	X	X	0	X	0	0	0	0
6	X	0	X	0	X	0	0	0	0
7	X	0	0	X	X	0	X	0	0

iterated AR(1) regression that $\hat{\Phi}_{1(1)}^{(1)} = .87$, which is quite close to the value of .92 in (4.2).

It was pointed out to us by the referees that an AR(7) model as suggested by inspection of the sample PACF would give a better fit for this data set, and the spectrum of the fitted AR(7) model would resemble more closely the sample spectrum obtained by using the Parzen window than that of the fitted ARMA(1, 1) model would. In fact, upon eliminating insignificant terms, a fitted AR model would be

$$(1 - .36B - .23B^2 - .23B^7)Z_t = 3.23 + a_t, \quad \hat{\sigma}_a^2 = .091. \quad (4.3)$$

Now, the behavior of the seventh column of Table 6 might suggest an ARMA(1, 7) model, and after simplification we would arrive at

$$(1 - .84B)Z_t = 2.73 + (1 - .50B + .19B^7)a_t, \quad \hat{\sigma}_a^2 = .094. \quad (4.4)$$

The two models (4.3) and (4.4) are essentially the same, as can be seen, for example, by expanding $(1 - .50B + .19B^7)^{-1}(1 - .84B)$ in (4.4), but the improvement obtained by adding an extra parameter to the ARMA(1, 1) model seems slight at best.

Example 4. Finally, consider the caffeine data analyzed by Hamilton and Watts (1978) and by Chatfield (1979). The data consist of 178 observations of caffeine levels in instant coffee taken every weekday. Thus a "seasonal" association of lag 5 is expected and in fact a model of the form

$$(1 - \phi B)(1 - B)Z_t = (1 - \theta B^5)a_t \quad (4.5)$$

has been suggested by these authors. Table 7 shows the ESACF and the corresponding indicator symbols for this set of data. A triangular cutting-off pattern shown in the sixth to the tenth columns, except for one X in the ninth column, clearly suggests an ARMA(2, 5) model for the original process. Note that the lower X in the ninth column is only slightly larger than the two standard deviation criterion used that is approximately equal to .16.

Table 7. The ESACF of the Caffeine Data

	MA									
AR	0	1	2	3	4	5	6	7	8	9
					a. The ESACF Table					
0	.89	.73	.58	.48	.42	.46	.49	.48	.42	.34
1	.30	.05	−.13	−.14	−.43	.07	.25	.25	.12	.02
2	.13	.08	−.11	.03	−.42	.15	.12	.13	.07	.01
3	−.27	.24	−.07	−.01	−.48	.33	−.08	.08	−.02	.08
4	−.37	−.20	−.02	.04	−.51	.30	.31	.01	−.05	.08
5	−.10	−.19	−.02	.40	−.45	.33	.06	−.05	−.18	−.02
					b. The Indicator Symbols					
0	X	X	X	X	X	X	X	X	X	X
1	X	0	0	0	X	0	X	X	0	0
2	0	0	0	0	X	0	0	0	0	0
3	X	X	0	0	X	X	0	0	0	0
4	X	X	0	0	X	X	X	0	0	0
5	0	X	0	X	X	X	0	0	X	0

All of the four examples given are nonseasonal in nature. The ESACF approach can be readily employed to help identify nonstationary seasonal models such as $U(B) = (1 - B)(1 - B^s)$. Examples of this kind will be discussed in a later article.

5. PROPERTIES OF THE ITERATED AR ESTIMATES

5.1 A Recursion

In this section, we show that the recursion (2.7) holds for the iterated AR estimates. Let

$$\hat{\Phi}_k^{(j)}(B) = 1 - \hat{\Phi}_{1(k)}^{(j)}B - \cdots - \hat{\Phi}_{k(k)}^{(j)}B^k \quad (5.1)$$

be the estimate of the AR polynomial in the jth iterated AR(k) regression (2.5).

Lemma 5.1. For any positive integers k and j, $\hat{e}_{k,t}^{(j)} = \hat{e}_{k+1,t}^{(j-1)}$.

Proof. Since $\hat{e}_{k,t-1}^{(0)} = \hat{\Phi}_k^{(0)}(B)Z_{t-1}$, it is clear from its definition in (2.5) that the first iterated AR(k) regression can be rewritten as an ordinary AR($k + 1$) fitting. Therefore, $\hat{e}_{k,t}^{(1)} = \hat{e}_{k+1,t}^{(0)}$. Repeating this argument, it is easily seen that $\hat{e}_{k,t}^{(j)} = \hat{e}_{k+j,t}^{(0)}$, and the lemma follows at once.

Lemma 5.2. For any positive integers k and j,

$$\hat{\Phi}_{k+j}^{(0)}(B) = \hat{\Phi}_k^{(j)}(B) - \sum_{i=1}^{j} \hat{\beta}_{i(k)}^{(j)} \hat{\Phi}_{k+j-i}^{(0)}(B)B^i.$$

Proof. This follows from the definition of $\hat{e}_{k,t}^{(j)}$ in (2.6) and $\hat{e}_{k,t}^{(j)} = \hat{e}_{k+j,t}^{(0)}$ as argued in Lemma 5.1.

Lemma 5.3. For any positive integers k and j,

$$\hat{\Phi}_k^{(j)}(B) = \hat{\Phi}_{k+1}^{(j-1)}(B) + \hat{\alpha}_k^{(j-1)}\hat{\Phi}_k^{(j-1)}(B)B,$$

where

$$\hat{\alpha}_k^{(j-1)} = -\hat{\Phi}_{k+1(k+1)}^{(j-1)}/\hat{\Phi}_{k(k)}^{(j-1)}.$$

Proof. This result is trivial for $j = 1$. For $j > 1$, by changing k to $k + 1$, and j to $j - 1$ in the equation of Lemma 5.2, we have that

$$\hat{\Phi}_{k+j}^{(0)}(B) = \hat{\Phi}_{k+1}^{(j-1)}(B) - \sum_{i=1}^{j-1} \hat{\beta}_{i(k+1)}^{(j-1)}\hat{\Phi}_{k+j-i}^{(0)}(B)B^i.$$

It follows by comparing this result with the equation of Lemma 5.2 that

$$\hat{\Phi}_k^{(j)}(B) = \hat{\Phi}_{k+1}^{(j-1)}(B) + \hat{\gamma}_{1(k)}^{(j)}\hat{\Phi}_{k+j-1}^{(0)}(B)B + \sum_{i=2}^{j-1} \hat{\gamma}_{i(k)}^{(j)}\hat{\Phi}_{k+j-i}^{(0)}(B)B^i + \hat{\beta}_{j(k)}^{(j)}\hat{\Phi}_k^{(0)}(B)B^j \quad (5.2)$$

where $\hat{\gamma}_{i(k)}^{(j)} = \hat{\beta}_{i(k)}^{(j)} - \hat{\beta}_{i(k+1)}^{(j-1)}$, $i = 1, \ldots, j - 1$.

Next, upon applying Lemma 5.2 to $\Phi_{k+j-1}^{(0)}(B)$, (5.2) becomes

$$\hat{\Phi}_k^{(j)}(B) - \hat{\Phi}_{k+1}^{(j-1)}(B) - \hat{\gamma}_{1(k)}^{(j)}\hat{\Phi}_k^{(j-1)}(B)B = \sum_{i=2}^{j-1}[\hat{\gamma}_{i(k)}^{(j)} - \hat{\gamma}_{1(k)}^{(j)}\hat{\beta}_{i-1(k)}^{(j-1)}]\hat{\Phi}_{k+j-i}^{(0)}(B)B^i + [\hat{\beta}_{j(k)}^{(j)} - \hat{\gamma}_{1(k)}^{(j)}\hat{B}_{j-1(k)}^{(j-1)}]\hat{\Phi}_k^{(0)}(B)B^j. \quad (5.3)$$

Notice that the left side of (5.3) is a polynomial in B of degree $k+1$. Therefore, those coefficients of B^{k+i}, $i \geq 2$, in the polynomial of the right side of (5.3) must be zero. Furthermore, for given n all the iterated LS estimates $\hat{\Phi}_{l(k)}^{(i)}$ are nonzero (see Remark following this proof.) Consequently,

$$\hat{\beta}_{j(k)}^{(j)} - \hat{\gamma}_{1(k)}^{(j)}\hat{\beta}_{j-1(k)}^{(j-1)} = 0,$$
$$\hat{\gamma}_{i(k)}^{(j)} - \hat{\gamma}_{1(k)}^{(j)}\hat{\beta}_{i-1(k)}^{(j-1)} = 0,$$

and then

$$\hat{\Phi}_k^{(j)}(B) = \hat{\Phi}_{k+1}^{(j-1)}(B) + \hat{\gamma}_{1(k)}^{(j)}\hat{\Phi}_k^{(j-1)}(B)B. \quad (5.4)$$

Finally, by equating the coefficients of B^{k+1} in (5.4), we have that

$$\hat{\gamma}_{1(k)}^{(j)}\hat{\Phi}_{k(k)}^{(j-1)} = -\hat{\Phi}_{k+1(k+1)}^{(j-1)}. \quad (5.5)$$

This completes the proof.

Remark. We may assume that all the $\hat{\Phi}_{l(k)}^{(i)}$ are nonzero. This follows from the assumption that the process Z_t is not purely deterministic so that all the iterated regression estimates are, as continuous random variables, nonzero with probability one for fixed n.

5.2 Relationship to Solutions of Moment Equations

Suppose that the process $\{Z_t\}$ is stationary. We consider the relationship between the iterated AR estimates and the solutions of sample moment equations (or the estimated generalized Yule-Walker equations). The normal equations of the jth iterated AR(k) regression (2.5) are

$$\sum Z_{t-h}e_{k,t}^{(j)} = 0, \quad h = 1, \ldots, k \quad (5.6)$$

$$\sum \hat{e}_{k,t-i}^{(j-i)}e_{k,t}^{(j)} = 0, \quad i = 1, \ldots, j, \quad (5.7)$$

where, and in what follows, the summation notation $\sum$ denotes, unless otherwise indicated, summing over t from $k+j+1$ to n. Since $\hat{e}_{k,t-i}^{(m)} = \hat{e}_{k+m,t-i}^{(0)} = \hat{\Phi}_{k+m}^{(0)}(B)Z_{t-i}$, it is easy to see that (5.6) and (5.7) reduce to

$$\sum Z_{t-h}e_{k,t}^{(j)} = 0, \quad h = 1, \ldots, k+j. \quad (5.8)$$

We can divide these $k+j$ equations in (5.8) into the following two systems of equations:

1. For $h = 1, \ldots, j$:

$$\mathbf{H}_j\boldsymbol{\beta}_j = \mathbf{V}_j \quad (5.9)$$

where $\boldsymbol{\beta}_j = (\beta_{1(k)}^{(j)}, \ldots, \beta_{j(k)}^{(j)})'$, $\mathbf{H}_j = \sum \mathbf{R}_{j,t}\mathbf{Z}_{j,t}'$ and $\mathbf{V}_j = \sum \mathbf{Z}_{j,t}W_{j,t}$ with $\mathbf{R}_{j,t} = (\hat{e}_{k,t-1}^{(j-1)}, \ldots, \hat{e}_{k,t-j}^{(0)})'$, $\mathbf{Z}_{j,t} = (Z_{t-1}, \ldots, Z_{t-j})'$, and $W_{j,t} = Z_t - \sum_{l=1}^{k}\Phi_{l(k)}^{(j)}Z_{t-l}$.

2. For $h = j+1, \ldots, k+j$:

$$\sum Z_{t-h}W_{j,t} = \sum_{i=1}^{j}\beta_{i(k)}^{(j)}[\sum Z_{t-h}\hat{e}_{k,t-i}^{(j-i)}]. \quad (5.10)$$

We now discuss some asymptotic properties of the solutions of (5.9) and (5.10). First, let $\mathbf{A}_h^{(v)}$ be the $h \times h$ matrix

$$\mathbf{A}_h^{(v)} = (\mathbf{D}_h^{(v)}, \mathbf{D}_h^{(v-1)}, \ldots, \mathbf{D}_h^{(v-h+1)}) \quad (5.11)$$

where $\mathbf{D}_h^{(l)} = (\rho_l, \rho_{l+1}, \ldots, \rho_{l+h-1})'$ and ρ_l is the autocorrelation of Z_t at lag l. Then we have the following property for the estimated residuals:

Lemma 5.4.

$$\sum_{k+1}^{n}\hat{e}_{k,t}^{(0)}Z_t \Big/ \sum_{1}^{n}Z_t^2 \xrightarrow{P} \det(\mathbf{A}_{k+1}^{(0)})/\det(\mathbf{A}_k^{(0)}).$$

Proof.

$$\sum_{k+1}^{n}\hat{e}_{k,t}^{(0)}Z_t \Big/ \sum_{1}^{n}Z_t^2 = \sum_{k+1}^{n}[\hat{\Phi}_k^{(0)}(B)Z_t]Z_t \Big/ \sum_{1}^{n}Z_t^2 \doteq 1 - \sum_{l=1}^{k}\hat{\Phi}_{l(k)}^{(0)}r_{l(0)}.$$

The result then follows directly from the column expansion of determinant and asymptotic covergence properties of sample ACF and PACF.

Next, recall that $\hat{e}_{k,t}^{(j-i)}$ is the estimated residual of the $(j-i)$th iterated AR(k) regression. By the least squares theory,

$$\sum_{t_0}^{n}Z_{t-h}\hat{e}_{k,t}^{(j-i)} = 0, \quad h = 1, \ldots, k, \quad (5.12)$$

where $t_0 = k+j-i+1$. Furthermore, since $\hat{e}_{k,t}^{(j-i)} = \hat{e}_{k+j-i,t}^{(0)}$ we also have that

$$\sum_{t_0}^{n}Z_{t-h}\hat{e}_{k,t}^{(j-i)} = 0, \quad h = k+1, \ldots, k+j-i. \quad (5.13)$$

From (5.12) and (5.13), one can easily see that

$$\sum Z_{t-h}\hat{e}_{k,t-i}^{(j-i)} = O_p(1), \quad h = i+1, \ldots, k+j \quad (5.14)$$

because each equation only contains a fixed finite number of terms independent of n.

Now consider (5.9). By Lemma 5.1 and Lemma 5.4, the diagonal elements of $\mathbf{H}_j$ can be normalized to yield

$$(\sum Z_t^2)^{-1}(\sum Z_{t-h}\hat{e}_{k,t-h}^{(j-h)}) = (\sum Z_t^2)^{-1}(\sum Z_{t-h}\hat{e}_{k+j-h,t-h}^{(0)}) \xrightarrow{P} \det(\mathbf{A}_{k+j-h+1}^{(0)})/\det(\mathbf{A}_{k+j-h}^{(0)}), \quad h = 1, \ldots, j. \quad (5.15)$$

By (5.14) and (5.15), $(\sum Z_t^2)^{-1}\mathbf{H}_j$ converges in probability to a nonsingular upper triangular matrix. Writing

$\hat{\boldsymbol{\beta}}_j' = (\hat{\beta}_{1(k)}^{(j)}, \ldots, \hat{\beta}_{j(k)}^{(j)})$, this implies that

$$\hat{\boldsymbol{\beta}}_j = [(\sum Z_t^2)^{-1}\mathbf{H}_j]^{-1}(\sum Z_t^2)^{-1}\hat{\mathbf{V}}_j \qquad (5.16)$$

always exists if $\hat{\mathbf{V}}_j$ does, where $\hat{\mathbf{V}}_j$ is obtained from $\mathbf{V}_j$ upon replacing $W_{j,t}$ by $\hat{W}_{j,t} = \hat{\Phi}_k^{(j)}(B)Z_t$. In other words, the existence of solutions of the normal equations (5.8) only depends on the existence of the $\hat{\Phi}_{l(k)}^{(j)}$'s.

We next consider the system of equations (5.10). If $\hat{\Phi}_{l(k)}^{(j)}$, $l = 1, \ldots, k$, are finite and satisfy these equations, then $\hat{\boldsymbol{\beta}}_j = O_p(1)$ by (5.16), and upon dividing through the equations by $\sum Z_t^2$ (5.10) becomes

$$r_h - \sum_{l=1}^{k} \hat{\Phi}_{l(k)}^{(j)} r_{h-l} = O_p(n^{-1}),$$

$$h = j + 1, \ldots, k + j. \qquad (5.17)$$

It follows from (5.16) and (5.17) that, if $\det(\mathbf{A}_k^{(j)}) \neq 0$, then asymptotically the normal equations (5.8) have a unique solution. Also, by ignoring the $O_p(n^{-1})$ term, we see that the equations in (5.17) are in fact the estimated generalized Yule-Walker equations considered recently by several authors for identification of ARMA models, for example, Woodward and Gray (1981), and Jenkins and Alavi (1981). Thus any solution $\{\hat{\Phi}_{l(k)}^{(j)}; l = 1, \ldots, k\}$ of the normal equations (5.8) is asymptotically equivalent to a solution of the estimated generalized Yule-Walker equations. We can summarize these results into the following lemma:

Lemma 5.5. Suppose Z_t follows a stationary ARMA model. If $\det(\mathbf{A}_k^{(j)}) \neq 0$, then asymptotically the LS estimates $\hat{\Phi}_{l(k)}^{(j)}$, $l = 1, \ldots, k$, for the jth iterated AR(k) regression are unique and are equivalent to the solution of the estimated generalized Yule-Walker equations.

5.3 Consistency Properties of Iterated AR Estimates

In this subsection we consider the consistency properties of the iterated AR estimates. For stationary ARMA processes, the consistency results are established through the estimated generalized Yule-Walker equations. However, this approach breaks down for the nonstationary situation, and we employ some properties of the OLS estimates of autoregressions shown in Tiao and Tsay (1983) to prove the results.

Stationary ARMA processes. For the stationary case ($\Phi(B) = \phi(B)$) we use the notation $\hat{\phi}_{l(k)}^{(j)}$ to denote the LS estimates of the jth iterated AR(k) regression and reserve $\hat{\Phi}_{l(k)}^{(j)}$ for the general models. First let us recall some known consistency properties of solutions of the estimated generalized Yule-Walker equations. Specifically, Gersch (1970) has shown that for stationary ARMA(p, q) models the solutions of

$$r_h - \sum_{l=1}^{p} \alpha_l r_{h-l} = 0, \quad h = q + 1, \ldots, q + p \qquad (5.18)$$

give consistent estimators of the AR parameters and the convergence rate is $O_p(n^{-1/2})$; that is,

$$\hat{\alpha}_l = \phi_l + O_p(n^{-1/2}), \quad l = 1, \ldots, p.$$

This result hinges on the fact that the equations in (5.18) converge in probability to the generalized Yule-Walker equations

$$\rho_h - \sum_{l=1}^{p} \phi_l \rho_{h-l} = 0, \quad h = q + 1, \ldots, q + p$$

and that the determinant $\det(\mathbf{A}_p^{(q)}) \neq 0$. To establish the consistency properties of the $\hat{\phi}_{l(k)}^{(j)}$'s, we need the following lemma:

Lemma 5.6. If Z_t follows a stationary ARMA(p, q) model, then (a) $\det(\mathbf{A}_p^{(j)}) \neq 0$ for $j \geq q$ and (b) $\det(\mathbf{A}_k^{(q)}) \neq 0$ for $k \geq p$.

Proof. We have that $\det(\mathbf{A}_p^{(j)}) = (-\phi_p)^{j-q} \det(\mathbf{A}_p^{(q)})$ for $j > q$ and $\det(\mathbf{A}_k^{(q)}) = [\phi(B)\rho_q]^{k-p} \det(\mathbf{A}_p^{(q)})$ for $k > p$ where the backshift operator B works on the subscript of ρ. The lemma then follows directly from $\det(\mathbf{A}_p^{(q)}) \neq 0$, $\phi_p \neq 0$, and $\phi(B)\rho_q \neq 0$.

From Lemma 5.6, one can readily show that for stationary ARMA(p, q) models the solutions of the equations

$$r_h - \sum_{l=1}^{k} \alpha_l r_{h-l} = 0, \quad h = j + 1, \ldots, j + k$$

for (i) $k \geq p$ and $j = q$, or (ii) $k = p$ and $j > q$ also give consistent estimates of the AR parameters. Using this result, (5.17) and Lemma 5.5 we immediately arrive at the following theorem:

Theorem 5.1. Suppose that Z_t is a stationary ARMA(p, q) process. Then, $\hat{\phi}_{l(k)}^{(j)} = \phi_l + O_p(n^{-1/2})$, $l = 1, \ldots, k$ if (i) $k \geq p$ and $j = q$ or (ii) $k = p$ and $j > q$, where it is understood that $\phi_l = 0$ for $l > p$.

Nonstationary ARMA processes. Suppose now that the ARMA process Z_t in (1.1) is nonstationary; that is, $U(B) \neq 1$. In this case, it is shown in our earlier article (Tiao and Tsay 1983) that the OLS estimates of an AR(k) regression with $k \geq d$ are asymptotically given by

$$\hat{\Phi}_k^{(0)}(B) \doteq \hat{U}_d^{(0)}(B)\hat{\phi}_m^{(0)}(B), \qquad (5.19)$$

where $m = k - d$ and the estimated polynomials $\hat{\Phi}_k^{(0)}(B)$, $\hat{U}_d^{(0)}(B)$, and $\hat{\phi}_m^{(0)}(B)$ are, respectively, the OLS estimates of the autoregressions

$$Z_t = \sum_{l=1}^{k} \Phi_{l(k)}^{(0)} Z_{t-l} + e_{k,t}^{(0)},$$

$$Z_t = \sum_{l=1}^{d} U_{l(d)}^{(0)} Z_{t-l} + e_{d,t}^{(0)},$$

and

$$W_t = \sum_{l=1}^{m} \phi_{l(m)}^{(0)} W_{t-l} + f_t^{(0)},$$

where $W_t = U(B)Z_t$ is a stationary process and follows the ARMA$(p - d, q)$ model $\phi(B)W_t = \theta(B)a_t$. The same article also shows that

$$\hat{U}_d^{(0)}(B) = U(B) + O_p(n^{-1}). \tag{5.20}$$

We return to the jth iterated AR(k) regression (2.5).

Lemma 5.7. $\hat{\Phi}_k^{(j)}(B) \doteq \hat{U}_d^{(0)}(B)\hat{\phi}_m^{(j)}(B), j \geq 0$ and $m = k - d$.

Proof. We prove this lemma by induction on j. Suppose that the lemma holds for $j = h$; that is, we assume that

$$\hat{\Phi}_k^{(h)}(B) \doteq \hat{U}_d^{(0)}(B)\hat{\phi}_m^{(h)}(B) \quad \text{for any } k \geq d. \tag{5.21}$$

By Lemma 5.3,

$$\hat{\Phi}_k^{(h+1)}(B) = \hat{\Phi}_{k+1}^{(h)}(B) + \hat{\alpha}\hat{\Phi}_k^{(h)}(B)B, \tag{5.22}$$

where $\hat{\alpha} = -\hat{\Phi}_{k+1(k+1)}^{(h)}/\hat{\Phi}_{k(k)}^{(h)}$. Next, by the induction assumption (5.21), we have that

$$\hat{\Phi}_{k+1}^{(h)}(B) \doteq \hat{U}_d^{(0)}(B)\hat{\phi}_{m+1}^{(h)}(B),$$
$$\hat{\Phi}_k^{(h)}(B) \doteq \hat{U}_d^{(0)}(B)\hat{\phi}_m^{(h)}(B). \tag{5.23}$$

It is then clear that

$$\hat{\alpha} = -\hat{\Phi}_{k+1(k+1)}^{(h)}/\hat{\Phi}_{k(k)}^{(h)} \doteq -\hat{\phi}_{m+1(m+1)}^{(h)}/\hat{\phi}_{m(m)}^{(h)}. \tag{5.24}$$

Therefore, by (5.23) and (5.24), Equation (5.22) becomes

$$\hat{\Phi}_k^{(h+1)}(B) \doteq \hat{U}_d^{(0)}(B)[\hat{\phi}_{m+1}^{(h)}(B) + \hat{\alpha}\hat{\phi}_m^{(h)}(B)B]$$
$$= \hat{U}_d^{(0)}(B)\hat{\phi}_m^{(h+1)}(B).$$

Thus the result also holds for $j = h + 1$. By (5.19), the lemma is true for $h = 0$, and hence the proof is complete.

Theorem 2.1 follows directly from the preceding lemma and (5.20). By Lemma 5.7 and (5.20), it is clear that $\hat{\Phi}_k^{(j)}(B)$ will be consistent for $\Phi(B)$ if and only if $\hat{\phi}_m^{(j)}(B)$ is for $\phi(B)$. Consequently, by Theorem 5.1, we have the general result given in Theorem 2.2. It is important to note that Theorem 2.2 covers both nonstationary and stationary ARMA processes. By contrast, it has been shown (see Quinn 1980 and Tiao and Tsay 1983) that the asymptotic behaviors of the sample ACF of a nonstationary process are dominated by those nonstationary roots with the highest multiplicity, and thus the estimated generalized Yule-Walker equations cannot always provide consistent estimates of the AR coefficients. Consequently, the asymptotic equivalence relationship (see Lemma 5.5) between the iterated AR estimates and the solutions of the estimated generalized Yule-Walker equations breaks down for the nonstationary processes, especially in the following situations: (i) when both $\phi(B)$ and $U(B)$ in (1.1) are not equal to 1; that is, when some zeros of $\Phi(B)$ lie outside and some are on the unit circle, and (ii) when $\Phi(B)$ has all its zeros on the unit circle but not all the zeros are distinct. For proofs, see for example Corollary 2.6 in our earlier article (Tiao and Tsay 1983).

6. PROPERTIES OF THE ESACF

We derive in this section the asymptotic properties of the extended sample autocorrelation function $r_{j(k)}$ stated in (3.5) and (3.6).

Proof of (3.5)

First, with respect to the "cutting-off" property (3.5), the key step is to show that $r_j(W_{k,t}^{(j)})$ is asymptotically equivalent to $r_j(W_{k,t})$ where $W_{k,t} = \Phi_k(B)Z_t$ for $k = p$.

Consider the ARMA(p, q) model of (1.1). If $U(B) \neq 1$, we can factor it into

$$U(B) = \prod_{i=1}^{m} U_i(B) \tag{6.1}$$

where $U_i(B) = 1 - \sum_{j=1}^{d_i} U_{j(i)}B^j$ such that (i) $\sum_{i=1}^{m} d_i = d$; (ii) $U_i(B)$ is a factor of $U_{i+1}(B)$; and (iii) all zeros of each $U_i(B)$ are simple; see Equations (3.1) and (3.2) of Tiao and Tsay (1983) for details. Then, defining

$$Z_{1,t} = Z_t,$$
$$Z_{i,t} = U_{i-1}(B)Z_{i-1,t}, \quad \text{for } i = 2, \ldots, m + 1, \tag{6.2}$$

the original process Z_t can be rewritten as

$$Z_t = \sum_{i=1}^{m}\sum_{v=1}^{d_i} U_{v(i)}Z_{i,t-v} + \sum_{h=1}^{p-d}\phi_h Z_{m+1,t-h} + e_t. \tag{6.3}$$

For (6.3), it can be shown (see Theorem 3.1 of Tiao and Tsay 1983) that

$$\hat{U}_{v(i)} = U_{v(i)} + O_p(n^{i-m+1}),$$
$$v = 1, \ldots, d_i; \quad i = 1, \ldots, m \tag{6.4}$$

where $\hat{U}_{v(i)}$ denotes the OLS estimates of $U_{v(i)}$ in the regression

$$Z_t = \sum_{i=1}^{m}\sum_{v=1}^{d_i} U_{v(i)}Z_{i,t-v} + W_t,$$

which, of course, is equivalent to the AR(d) fitting on Z_t. Furthermore, from the transformation (6.2), the OLS estimates of the AR(d) fitting on Z_t are given by

$$\hat{U}_d^{(0)}(B) = 1 - \sum_{i=1}^{m}\left[\sum_{v=1}^{d_i}\left\{\prod_{k=1}^{i-1} U_k(B)\right\}\hat{U}_{v(i)}B^v\right]. \tag{6.5}$$

It follows from Lemma 2.5 in Tiao and Tsay (1983), or from Equations (3.7) and (3.8) there, that

$$\sum^{n} Z_{i,t}Z_{j,t} = O_p(n^{2m-i-j+2}) \qquad i,j = 1, \ldots, m,$$
$$\sum^{n} Z_{i,t}Z_{m+1,t} = O_p(n^{m-i+1}) \qquad i = 1, \ldots, m. \tag{6.6}$$

Now consider the jth iterated AR(p) regression of Z_t for $j \geq q$. By Lemma 5.7, the transformation (6.2) and the Equation (4.17) of our earlier article (1983),

$$\hat{\Phi}_p^{(j)}(B) \doteq \hat{U}_d^{(0)}(B) - U(B)\sum_{h=1}^{p-d}\hat{\phi}_h^{(j)}B^h. \tag{6.7}$$

Let $r_s(X_t)$ be the sample ACF of X_t at lag s. Then we have the following result:

Lemma 6.1. If Z_t is an ARMA(p, q) process, then $r_s(\hat{e}_t^{(j)})$ is asymptotically equivalent to $r_s(e_t)$ where $\hat{e}_t^{(j)} = \hat{\Phi}_p^{(j)}(B)Z_t$.

Proof. By (6.7), (6.5), and (6.2),

$$\hat{e}_t^{(j)} = \hat{\Phi}_p^{(j)}(B)Z_t = Z_t - \sum_{i=1}^{m}\sum_{v=1}^{d_i} \hat{U}_{v(i)}Z_{i,t-v} - \sum_{h=1}^{p-d} \hat{\phi}_h^{(j)}Z_{m+1,t-h}.$$

Substituting (6.3) for Z_t, we have that

$$\hat{e}_t^{(j)} = e_t - \sum_{i=1}^{m}\sum_{v=1}^{d_i} (\hat{U}_{v(i)} - U_{v(i)})Z_{i,t-v} - \sum_{h=1}^{p-d} (\hat{\phi}_h^{(j)} - \phi_h)Z_{m+1,t-h}.$$

Next, by (6.6), (6.4), $\sum^n Z_{m+1,t}^2 = O_p(n)$ and $\sum^n e_t^2 = O_p(n)$, it is easily seen that

$$\sum^n \hat{e}_t^{(j)}\hat{e}_{t-s}^{(j)} = \sum^n e_t e_{t-s}[1 + O_p(n^{-1/2})].$$

Therefore, the result holds.

The implication of Lemma 6.1 is that we can treat the consistent estimates of AR parameters as their corresponding true values in discussing the asymptotic properties of the SACF of the transformed series. Equation (3.5) follows from this lemma.

The Overfitting Result (3.6)

In what follows we say that the estimated polynomial $\hat{\Phi}_k^{(j)}(B) \xrightarrow{P} G(B)$ where $G(B)$ is some polynomial in B if all the coefficient estimates in $\hat{\Phi}_k^{(j)}(B)$ converge in probability to the corresponding coefficients in $G(B)$. From the recursion formula in Lemma 5.3, we change j to $q + i$ and put $k = p + h$ so that

$$\hat{\Phi}_{p+h}^{(q+i)}(B) = \hat{\Phi}_{p+h+1}^{(q+i-1)}(B) + \hat{\alpha}_h^{(i-1)}\hat{\Phi}_{p+h}^{(q+i-1)}(B)B, \quad (6.8)$$

where

$$\hat{\alpha}_h^{(i-1)} = -\hat{\Phi}_{p+h+1(p+h+1)}^{(q+i-1)}/\hat{\Phi}_{p+h(p+h)}^{(q+i-1)}. \quad (6.9)$$

For ARMA(p, q) models, we consider the asymptotic behavior of $\hat{\Phi}_{p+h}^{(q+i)}(B)$ for $i \geq 1$ and $h \geq 1$. To motivate the discussion, consider first the case $i = 1$. From Theorem 2.2,

$$\hat{\Phi}_{p+h}^{(q)}(B) \xrightarrow{P} \Phi(B) \quad \text{for } h \geq 0, \quad (6.10)$$

and

$$\hat{\Phi}_{p+l(p+h)}^{(q)} = O_p(n^{-1/2}) \quad \text{for } l = 1, \ldots, h;\ h \geq 1. \quad (6.11)$$

Thus in (6.8) with $i = 1$, for $h \geq 1$,

$$\hat{\alpha}_h^{(0)} \xrightarrow{P} \alpha_h^{(0)}$$

and

$$\hat{\Phi}_{p+h}^{(q+1)}(B) \xrightarrow{P} \Phi(B)(1 + \alpha_h^{(0)}B), \quad (6.12)$$

where $\alpha_h^{(0)}$ is some continuous real-valued random variable, and hence it is nonzero with probability one. On the other hand, by Theorem 2.2 directly or the fact that $\hat{\Phi}_{p(p)}^{(q)} \xrightarrow{P} \Phi_p$ we have that, for $h = 0$,

$$\hat{\alpha}_0^{(0)} \xrightarrow{P} 0$$

and

$$\hat{\Phi}_p^{(q+1)}(B) \xrightarrow{P} \Phi(B). \quad (6.13)$$

The implication of the results in (6.12) and (6.13) on the ESACF is as follows. With $W_{p+h,t}^{(q+1)} = \hat{\Phi}_{p+h}^{(q+1)}(B)Z_t$, we have that, for large n,

$$W_{p+h,t}^{(q+1)} \doteq \theta(B)a_t, \quad h = 0$$
$$\doteq (1 + \alpha_h^{(0)}B)\theta(B)a_t, \quad h > 0. \quad (6.14)$$

It is easy to see that

$$r_{q+1(p+h)} \doteq 0, \quad h = 0$$
$$\doteq c(h, 1), \quad h > 0, \quad (6.15)$$

where $c(h, 1)$ is some arbitrary nonzero constant between -1 and 1 with probability one. By comparing (6.15) with (3.5) we see that while $r_{q+i(p)}$, $i > 0$, possess the "cutting-off" property similar to the SACF of an MA(q) model, this property is no longer true for $r_{q+i(p+h)}$ when more than p autoregressive parameters are included in the iterated regression. We now extend the result (6.12) to the general case $i \geq 1$.

We first define the polynomial function $H_h^{(i)}(B)$, $i \geq 0$, $h \geq i$ as

$$H_h^{(i)}(B) = H_{h+1}^{(i-1)}(B) + \alpha_h^{(i-1)}H_h^{(i-1)}(B)B, \quad i > 0 \quad (6.16)$$

and $H_h^{(0)}(B) = 1$, where $\alpha_h^{(i-1)}$ is a continuous real-valued random variable. Note that $H_h^{(i)}(B)$ is a polynomial of degree i and the coefficient of B^i is $\prod_{v=1}^{i} \alpha_h^{(i-v)}$.

Lemma 6.2. Suppose Z_t follows an ARMA(p, q) model. Then for $i \geq 1$ and $h \geq i$

$$\hat{\Phi}_{p+h}^{(q+i)}(B) \xrightarrow{P} \Phi(B)H_h^{(i)}(B). \quad (6.17)$$

Proof. We prove the result by induction on i. From (6.12), the lemma is true for $i = 1$. Suppose (6.17) holds for $i = b \geq 1$. Then, from (6.8),

$$\hat{\Phi}_{p+h}^{(q+b+1)}(B) = \hat{\Phi}_{p+h+1}^{(q+b)}(B) + \hat{\alpha}_h^{(b)}\hat{\Phi}_{p+h}^{(q+b)}(B)B. \quad (6.18)$$

By assumption,

$$\hat{\Phi}_{p+h+1}^{(q+b)}(B) \xrightarrow{P} \Phi(B)H_{h+1}^{(b)}(B)$$

and

$$\hat{\Phi}_{p+h}^{(q+b)}(B) \xrightarrow{P} \Phi(B)H_h^{(b)}(B), \qquad (6.19)$$

which imply that for $h > b$,

$$\hat{\Phi}_{p+h(p+h+1)}^{(q+b)} \xrightarrow{P} 0,$$
$$\hat{\Phi}_{p+h+1(p+h+1)}^{(q+b)} \xrightarrow{P} 0, \qquad (6.20)$$

and it can be shown by using the result (6.11) that the convergence rates in (6.20) are the same. It follows from (6.20) that $\hat{\alpha}_h^{(b)}$ converges in probability to a random variable, $\alpha_h^{(b)}$ say, which together with (6.8) and (6.19) imply that

$$\hat{\Phi}_{p+h}^{(q+b+1)}(B) \xrightarrow{P} \Phi(B)H_h^{(b+1)}(B) \quad \text{for } h \geq b+1,$$

and the lemma is proved.

The asymptotic behavior of $\hat{\Phi}_{p+h}^{(q+i)}(B)$ for $h < i$ is given by the following lemma:

Lemma 6.3. Suppose Z_t follows an ARMA(p, q) model. Then, for $h \geq 0$ and $l = 0, 1, 2, \ldots,$

$$\hat{\Phi}_{p+h}^{(q+h+l)}(B) \xrightarrow{P} \Phi(B)H_{h+l}^{(h)}(B).$$

Proof. We prove the result by induction on l. The result for $l = 0$ is proved in Lemma 6.2 and that for $l = 1$ can be easily shown. Now suppose the lemma is true for $l = 1, \ldots, b$. For $l = b + 1$

$$\hat{\Phi}_{p+h}^{(q+h+b+1)}(B) = \hat{\Phi}_{p+h+1}^{(q+h+b)}(B) + \hat{\alpha}_h^{(h+b)}\hat{\Phi}_{p+h}^{(q+h+b)}(B)B \qquad (6.21)$$

where

$$\hat{\alpha}_h^{(h+b)} = -\hat{\Phi}_{p+h+1(p+h+1)}^{(q+h+b)} / \hat{\Phi}_{p+h(p+h)}^{(q+h+b)}.$$

By assumption

$$\hat{\Phi}_{p+h+1}^{(q+h+b)}(B) \xrightarrow{P} \Phi(B)H_{h+b}^{(h+1)}(B)$$

and

$$\hat{\Phi}_{p+h}^{(q+h+b)}(B) \xrightarrow{P} \Phi(B)H_{h+b}^{(h)}(B). \qquad (6.22)$$

Thus,

$$\hat{\Phi}_{p+h+1(p+h+1)}^{(q+h+b)} \xrightarrow{P} \Phi_p \prod_{v=1}^{h+1} \alpha_{h+b}^{(h+1-v)}$$

and

$$\hat{\Phi}_{p+h(p+h)}^{(q+h+b)} \xrightarrow{P} \Phi_p \prod_{v=1}^{h} \alpha_{h+b}^{(h-v)}.$$

So that in (6.21),

$$\hat{\alpha}_h^{(h+b)} \xrightarrow{P} -\alpha_{h+b}^{(h)}. \qquad (6.23)$$

It follows from (6.22) and (6.23) that

$$\hat{\Phi}_{p+h}^{(q+h+b+1)}(B) \xrightarrow{P} \Phi(B) \times [H_{h+b}^{(h+1)}(B) - \alpha_{h+b}^{(h)} H_{h+b}^{(h)}(B)] = \Phi(B)H_{h+b+1}^{(h)}(B).$$

Hence the result also holds for $l = b + 1$, and the lemma follows.

We can now summarize the results in Theorem 2.2, Lemma 6.2, and Lemma 6.3 into the following theorem:

Theorem 6.1. Suppose Z_t follows an ARMA(p, q) model. Then, for $j \geq q$ and $k \geq p$, the jth iterated AR(k) polynomial $\hat{\Phi}_k^{(j)}(B)$ converges in probability as follows:

$$\hat{\Phi}_k^{(j)}(B) \xrightarrow{P} \Phi(B)\, H_{k-p}^{(j-q)}(B), \quad j - q \leq k - p$$
$$\xrightarrow{P} \Phi(B)\, H_{j-q}^{(k-p)}(B), \quad j - q > k - p$$

where the $H(B)$'s are defined in (6.16).

Corollary 6.1. Let $W_{k,t}^{(j)} = \hat{\Phi}_k^{(j)}(B)Z_t$. Then for sufficiently large n,

$$W_{k,t}^{(j)} \doteq \theta(B)\, H_{k-p}^{(j-q)}(B)a_t, \quad 0 \leq j - q \leq k - p$$
$$\doteq \theta(B)\, H_{j-q}^{(k-p)}(B)a_t, \quad j - q > k - p \geq 0,$$

where $H_h^{(i)}(B)$ is a polynomial of degree i.

This corollary clearly implies that the effect of overfitting stated in Section 3 holds for the ARMA models. One can now easily verify the property stated in (3.6).

7. DISCUSSIONS AND COMPARISONS

In this section some further properties of the ESACF approach are first briefly discussed. Then some further comparisons are made between this approach and other identification procedures in the literature.

For stationary processes, the ESACF can be interpreted in the following manner. The moment equations of a stationary ARMA(p, q) model can be written as

$$w_h = \rho_h - \sum_{l=1}^{p} \phi_l \rho_{h-l}, \quad h \geq 0,$$

where w_h are functions of the parameters ϕ's, θ's, and σ_a^2 such that

$$w_h \neq 0, \quad h = q$$
$$= 0, \quad h > q. \qquad (7.1)$$

Since the true value of q is unknown, one can in practice estimate the w_h's in (7.1) by solving the sample equations

$$w_h = r_h - \sum_{l=1}^{p} \phi_l r_{h-l}, \quad h = j$$

and

$$0 = r_h - \sum_{l=1}^{p} \phi_l r_{h-l}, \quad h = j+1, \ldots, j+p \qquad (7.2)$$

for $j = 1, 2, \ldots$. Denoting the estimate of w_h in (7.2) by $\hat{w}_{j(p)}$, one can then easily see from the definition of the ESACF that

$$r_{j(p)} \propto w_{j(p)}.$$

Thus for stationary ARMA models the ESACF can be

regarded as one method to estimate the cutting-off property (7.2).

The ESACF table only provides information on the "maximum" values of p and q, mainly because the ESACF is not the SACF of any particular transformed series of Z_t. For instance, consider Example 4 of Section 4. Although the ESACF seems to suggest that only lag 5 is important in the MA part as the model (4.5) indicates, we can only conclude an ARMA(2, 5) model for the data rather than an ARMA(2, 5) with MA factor $(1 - \theta B^5)$. In practice, one can either leave this problem to the estimation stage of the analysis or check the SACF of the corresponding transformed series $W_{2,t}^{(5)} = \hat{\Phi}_2^{(5)}(B)Z_t$.

As demonstrated by the examples in Section 4, the ESACF approach does not require differencing of data for nonstationary series. If one is interested in finding the nonstationary part of $\Phi(B)$, it can be obtained by ascertaining the iterated AR estimates once the values of p and q are specified (see, e.g., Example 2). In general, the iterated AR(d) estimates give the nonstationary factor $U(B)$ of (1.1).

In summary, once the orders of (p, q) are tentatively identified from the ESACF table, it seems advisable to study the nature of the corresponding estimated AR polynomial $\hat{\Phi}_p^{(q)}(B)$ and the SACF of $W_{p,t}^{(q)} = \hat{\Phi}_p^{(q)}(B)Z_t$ for parsimonious parameterization before going into a full-fledged maximum likelihood estimation. In particular, because of the faster convergence rate of the estimates of $U(B)$, nonstationary factors can usually be readily spotted and factored out of $\hat{\Phi}_p^{(q)}(B)$, leaving only $\phi(B)$ to be estimated along with $\theta(B)$.

The triangular cutting-off characteristic of the ESACF table in practice may become a rectangular or trapezoidal shape. In those cases, we can specify the model from the upper left vertex of the cutting-off pattern based on the preference for simple models. In fact, instead of the definition (3.4) one can define the ESACF as

$$r_{j(k)}^* = r_j(W_{k,t}^{(j-1)}).$$

Then, by Corollary 6.1, the cutting-off pattern of the ESACF will asymptotically become a rectangle. Experiences, however, show that (3.4) is a better tool for model identification.

We note that although in (1.1) we have assumed, as is usually done, that the a_t's are normally distributed, this assumption is in fact not crucial for the ESACF approach. For the consistency property of the ESACF to hold, all that is needed is that the a_t's are independently and identically distributed continuous random variables with finite fourth moments.

We now compare the ESACF with other approaches for model identification. For nonstationary models, the relative advantages of the present approach have already been mentioned in Section 4 along with the examples. In what follows, we restrict the comparisons within the stationary models, because most other approaches have been developed on the basis of the stationarity assumption.

ESACF Versus *S*-Array

For stationary processes, let the system of the jth estimated generalized Yule-Walker equations with order k be written as

$$\mathbf{A}(j, k)\hat{\boldsymbol{\phi}}(j, k) = \mathbf{G}(j + 1, k) \tag{7.3}$$

where $\mathbf{G}(j + 1, k) = (r_{j+1}, \ldots, r_{j+k})'$, $\mathbf{A}(j, k) = (\mathbf{G}'(j, k), \ldots, \mathbf{G}'(j - k + 1, k))$, and $\hat{\boldsymbol{\phi}}(j, k)$ denotes a vector of solutions. Then, it is well known that for stationary ARMA(p, q) models

$$\hat{\boldsymbol{\phi}}(j, k) \xrightarrow{P} \boldsymbol{\phi}(k) \quad \text{for } j \geq q \quad \text{and } k = p, \tag{7.4}$$

where $\boldsymbol{\phi}(k) = (\phi_1, \ldots, \phi_k)'$ is the vector of true AR coefficients. Therefore,

$$\hat{S}(j, p) \xrightarrow{P} 1 - \sum_{l=1}^{p} \phi_l \quad \text{for } j \geq q, \tag{7.5}$$

where $\hat{S}(j, p) = 1 - \mathbf{1}'(p)\hat{\boldsymbol{\phi}}(j, p)$ and $\mathbf{1}(p)$ is a p-dimension vector of ones. From the definition of the S-array (see, e.g., Gray, Kelly and McIntire 1978), it is easily seen that the basic idea of the S-array and its associates is to employ the consistency result (7.5). Recall from Section 5 that the iterated AR estimates are asymptotically equivalent to the solutions of the estimated generalized Yule-Walker equations for the stationary models. Thus the ESACF based on the iterated AR estimates clearly also hinges on the consistency property (7.4). Consequently, both approaches use the same consistency property but explore it in different manners. The main differences can be summarized as follows: (a) The asymptotic properties of the ESACF are available while those of the S-array are only partially known; (b) the present ESACF approach is much simpler to use in practice than the S-array approach, which usually requires examining several complex tables; (c) the present approach also provides directly initial AR estimates for further estimation while the S-array approach requires solving some equations to get them; and (d) the ESACF approach can efficiently resolve the overfitting problems while the S-array cannot.

ESACF Versus GPACF

Woodward and Gray (1981), Tiao and Box (1981), and Jenkins and Alavi (1981) propose to employ the last estimated coefficient of Equation (7.3) as a generalization of the sample PACF for order determination of the mixed models. This may be called the GPACF approach. Although both the ESACF and GPACF share the same feature of providing information on maximum order, they are different in several ways. First, the true value of the GPACF is no longer between -1 and 1. This not only causes confusion with the random terms due to overfittings but also makes the sampling properties of the GPACF unattractive. Consequently, it reduces the practical usefulness of the GPACF. Second, based on our simulation experience, the consistency property of the sample GPACF is much less stable than that of the

ESACF. Finally, it is hard to interpret the GPACF in the nonstationary case, especially when nonstationary factors have multiplicities greater than one.

Finally, Hannan and Rissanen (1982) have recently proposed a recursive procedure for the order determination in stationary ARMA models. However, there are several differences between their procedure and the ESACF approach. For instance, their second stage requires some special attention in order to avoid the difficulties caused by overfitting that is not required by the present approach; also, the ESACF approach can directly handle the nonstationary series.

[*Received July 1982. Revised August 1983.*]

REFERENCES

AKAIKE, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716–723.

BEGUIN, J.M., GOURIEROUX, C., and MONFORT, A. (1980), "Identification of a Mixed Autoregressive-moving Average Process: The Corner Method," in *Time Series*, ed. O.D. Anderson, Amsterdam: North-Holland, 423–436.

BOX, G.E.P., and JENKINS, G.M. (1976), *Time Series Analysis-Forecasting and Control*, San Francisco: Holden-Day.

CHATFIELD, C. (1979), "Inverse Autocorrelation," *Journal of the Royal Statistical Society*, Ser. A, 142, 363–377.

DURBIN, J. (1960), "The Fitting of Time Series Models," *Review of the International Statistical Institute*, 28, 233–244.

GERSCH, W. (1970), "Estimation of the Autoregressive Parameters of a Mixed Autoregressive Moving Average Time Series," *IEEE Transactions on Automatic Control*, AC-14, 583–588.

GRAY, H.L., KELLY, G.D., and MCINTIRE, D.D. (1978), "A New Approach to ARMA Modeling," *Communications in Statistics*, B7, 1–77.

HAMILTON, D.C., and WATTS, D.G. (1978), "Interpreting Partial Autocorrelation Functions of Seasonal Time Series Models," *Biometrika*, 65, 135–140.

HANNAN, E.J., and RISSANEN, J. (1982), "Recursive Estimation of Mixed Autoregressive Moving Average Order," *Biometrika*, 69, 81–94.

JENKINS, G.M., and ALAVI, A.S. (1981), "Some Aspects of Modeling and Forecasting Multivariate Time Series," *Journal of Time Series Analysis*, 2, 1–47.

KASHYAP, R.L. (1980), "Inconsistency of the AIC rule for Estimating the Order of Autoregressive Models," *IEEE Transactions on Automatic Control*, AC-25, 996–998.

MANN, H.B., and WALD, A. (1943), "On the Statistical Treatment of Linear Stochastic Difference Equations," *Econometrica*, 11, 173–220.

QUINN, B.G. (1980), "Limiting Behavior of Autocorrelation Function of ARMA Processes as Several Roots of Characteristic Equation Approach Unit Circle," *Communication in Statistics*, B9, 195–198.

SCHWARZ, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

SHIBATA, R. (1976), "Selection of the Order of an Autoregressive Model by Akaike's Information Criterion," *Biometrika*, 63, 117–126.

——— (1980), "Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process," *Annals of Statistics*, 8, 147–164.

TANIGUCHI, M. (1980), "On Selection of the Order of the Spectral Density Model for a Stationary Process," *Annals of the Institute of Statistical Mathematics*, 32, 401–419.

TIAO, G.C., and BOX, G.E.P. (1981), "Modeling Multiple Time Series With Applications," *Journal of the American Statistical Association*, 76, 802–816.

TIAO, G.C., and TSAY, R.S. (1983), "Consistency Properties of Least Squares Estimates of Autoregressive Parameters in ARMA Models," *Annals of Statistics*, 11, 856–871.

WHITTLE, P. (1963), "On the Fitting of Multivariate Autoregressions, and the Approximate Canonical Factorization of a Spectral Density Matrix," *Biometrika*, 50, 129–134.

WOODWARD, W.A., and GRAY, H.L. (1981), "On the Relationship Between the *S*-array and the Box-Jenkins Method of ARMA Model Identification," *Journal of American Statistical Association*, 76, 579–587.