AD-A091 515 SOUTHERN METHODIST UNIV DALLAS TEX DEPT OF STATISTICS F/G 12/1
MODELING SEASONAL ARMA PROCESSES.(U)
OCT 80 J D HART, H L GRAY N00014-75-C-0439
UNCLASSIFIED TR-140 NL

1 OF 1
AD A
091515

END
DATE
FILMED
12-80
DTIC

(4)

4

MODELING SEASONAL ARMA PROCESSES

by

Jeffrey D./Hart
H. L./Gray

Technical Rep[redacted] No. 140
Department of Statistics ONR Contract

Oct[redacted]80

Research sponsored by the Office of Naval Research
Contract N00014-75-C-0439
Project NR 042-280

TR-140

DIVISION OF MATHEMATICAL SCIENCES
Department of Statistic
Southern Methodist University
Dallas, Texas 75275

404198

MODELING SEASONAL ARMA PROCESSES

Jeffrey D. Hart and H. L. Gray
Southern Methodist University

INTRODUCTION

Gray, Kelley, and McIntire (1978) have introduced a method, based on arrays of numbers called R- and S-arrays, for identifying p and q in an ARMA(p,q) process. In addition, they have illustrated how the same method is useful in detecting nonstationary factors in an observed process, and in suggesting an appropriate transformation to stationarity. In the present paper special attention is given to the problem of modeling seasonal ARMA processes using the S-array method. A general definition is given for a seasonal process, and the procedure for identifying and modeling such processes is discussed in detail. Additionally, an interesting theorem characterizing the S-arrays (based upon the sample autocorrelation) of seasonal processes is stated and a proof indicated. Finally, a data set (the international airline data) which exhibits the properties of a seasonal process is analyzed using the method discussed, and two models for the data are proposed.

I. Definitions and Theorems

The following definitions and theorems provide the motivation for the remainder of the paper.

Definition 1. A stochastic process $\{X_t\}$, $t = 0, \pm 1, \pm 2, \ldots$ is said to be autoregressive of order p and moving average of order q, or ARMA(p,q), if

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + Z_t - \sum_{k=1}^{q} \theta_k Z_{t-k} , \qquad (1)$$

where the ϕ_k and θ_k are constants and $\{Z_t\}$ is a white noise process with finite variance. If we define the operator B by $BX_t = X_{t-1}$, then (1) may be written as

$$\phi(B)X_t = \theta(B)Z_t \qquad (2)$$

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

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

It is well known that $\{X_t\}$ is stationary if and only if all of the roots of $\phi(x)=0$ lie outside the unit circle. For our purposes, a nonstationary process will be one for which one or more of the roots of $\phi(x)=0$ lie on the unit circle.

Definition 2. Suppose $\{X_t\}$ is an ARMA(p,q) process with $\phi(B)X_t = \theta(B)Z_t$. A factor $\alpha(B)$ of $\phi(B)$ will be called a seasonal factor if

$$\alpha(B) = 1 - \alpha_1 B + B^2 , \ |\alpha_1| < 2 \qquad \text{or}$$

$$\alpha(B) = 1 + B ;$$

i.e. a factor $\alpha(B)$ is seasonal if the algebraic equation $\alpha(x) = 0$ has complex roots on the unit circle or the root -1.

Definition 3. An ARMA(p,q) process will be referred to as a seasonal process if it has one or more seasonal factors.

Definition 4. Let m be an integer and f be a real valued function. Further, let $f_m = f(m)$,

$$H_n[f_m] = \begin{vmatrix} f_m & f_{m+1} & \cdots & f_{m+n-1} \\ f_{m+1} & f_{m+2} & \cdots & f_{m+n} \\ \vdots & \vdots & & \\ f_{m+n-1} & f_{m+n} & \cdots & f_{m+2n-2} \end{vmatrix} ,$$

$H_0[f_m] \equiv 1$, and

$$H_{n+1}[1;\, f_m] = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ f_m & f_{m+1} & \cdots & f_{m+n} \\ f_{m+1} & f_{m+2} & \cdots & f_{m+n+1} \\ \vdots & \vdots & & \vdots \\ f_{m+n-1} & f_{m+n} & \cdots & f_{m+2n-1} \end{vmatrix} .$$

Now define

$$S_n(f_m) = \frac{H_{n+1}[1;\, f_m]}{H_n[f_m]} .$$

The S-array (for the function f) is the following array of numbers:

$m \setminus n$	1	2	...	k
$-\ell$	$S_1(f_{-\ell})$			
$-\ell+1$	$S_1(f_{-\ell+1})$	$S_2(f_{-\ell})$		
$\vdots$		$S_2(f_{-\ell+1})$		
	$\vdots$	$\vdots$		
$-\ell+k-1$			...	$S_k(f_{-\ell})$
$\vdots$				$\vdots$
-1	$S_1(f_{-1})$	$S_2(f_{-2})$	...	$S_k(f_{-k})$
- - -	- - -	- - -	- - -	- - -
0	$S_1(f_0)$	$S_2(f_{-1})$	...	$S_k(f_{-k+1})$
1	$S_1(f_1)$	$S_2(f_0)$		$S_k(f_{-k+2})$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$j-k+1$				$S_k(f_{j-2k+2})$
$\vdots$				
$j-1$	$S_1(f_{j-1})$	$S_2(f_{j-2})$		
j	$S_1(f_j)$			

A recursion relationship for calculating S-array values and a complete discussion of how S-arrays may be utilized in identifying p and q for an ARMA(p,q) process may be found in Gray, Kelley, and McIntire (1978).

Definition 5. Denote by $\rho_m(\lambda_1, \ldots, \lambda_p)$ the autocorrelation function of a stationary ARMA(p,q) process whose characteristic equation has roots $\lambda_1, \lambda_2, \ldots, \lambda_p$. Let $r_1, r_2, \ldots, r_p$ be the roots of the characteristic equation of an ARMA(p,q) process with j roots on the unit circle, which for convenience are denoted by $r_1, r_2, \ldots, r_j$. Now, define

$$\rho_m^* = \rho_m(r_1,r_2,\ldots,r_p) \;, \quad j = 0$$

and

$$\rho_m^* = \lim_{\alpha \to 1^+} \rho_m(\alpha r_1,\alpha r_2,\ldots,\alpha r_j,r_{j+1},\ldots,r_p), \; j = 1,2,\ldots,p \;.$$

Theorem 1. Suppose an ARMA(p+j,q) process has j roots of its characteristic equation on the unit circle and that M of these j roots are of highest multiplicity. Let $r_1,r_2,\ldots,r_M$ denote the M distinct roots on the unit circle which are of highest multiplicity. Then ρ_m^* satisfies a linear homogeneous difference equation of order M whose characteristic equation is

$$(1-r_1^{-1}x)(1-r_2^{-1}x) \;\ldots\; (1-r_M^{-1}x) = 0 \;.$$

For a proof of this theorem see Findley (1978) or Quinn (1980). The importance of Theorem 1 lies in its suggestion of the probable behavior of the sample autocorrelation function from nonstationary ARMA processes. The next two theorems will illustrate how S-arrays may be used to take advantage of the information in Theorem 1.

For the remainder of this paper we shall define the autocorrelation of a nonstationary process as ρ_m^* and hence will write $\rho_m \equiv \rho_m^*$.

<u>Theorem 2</u>. Let f_m denote either ρ_m or $(-1)^m \rho_m$ where ρ_m is the autocorrelation function of an ARMA(p,q) process. Then the condition that $S_\ell(f_m)$ is constant (as a function of m) for some $\ell \leq p$ is both necessary and sufficient for nonstationarity. Further, $\ell = M$ (where M is the same as that in Theorem 1) and

$$S_M(f_m) \equiv c_f = (-1)^M [1 - \sum_{k=1}^{M} a_{f,k} \beta_k] \quad \text{where}$$

$$\rho_m - \beta_1 \rho_{m-1} - \dots - \beta_M \rho_{m-M} = 0 \quad \text{for} \quad m > 0 \quad \text{and}$$

$$a_{f,k} = \begin{cases} 1 \;, & f_m \equiv \rho_m \\ (-1)^k, & f_m \equiv (-1)^m \rho_m \end{cases} .$$

The proof of this theorem follows from the proof of theroem 9 in Gray, Kelley, and McIntire (1978) and from the proof of the previous theorem in this paper.

<u>Theorem 3</u>. Suppose $\{X_t\}$ is an ARMA(p.q) process and that

$$\hat{\rho}_N(k) = \frac{\sum_{t=1}^{N-k} \left(X_t - \bar{X}_N\right)\left(X_{t+k} - \bar{X}_N\right)}{\sum_{t=1}^{N} \left(X_t - \bar{X}_N\right)^2} .$$

Then $\{X_t\}$ is nonstationary if and only if $\text{p-lim}_{N\to\infty} S_m(\hat{\rho}_N(k)) \equiv c$ for some $m \leq p$. Moreover, if $S_m(\rho_k)$ is defined for all k, then

$$c = (-1)^m (1 - \sum_{k=1}^{m} \beta_k), \quad \text{where}$$

$$\rho_k - \beta_1 \rho_{k-1} - \dots - \beta_m \rho_{k-m} = 0 .$$

This theorem has been proven by Findley in personal communications to H.L. Gray and by Morton (1980).

Theorems 1, 2, and 3 provide the basis for modeling seasonal data. Theorems 1 and 2 indicate respectively the effect which seasonal factors have on the autocorrelation function, and the way in which seasonal factors are manifested in S-arrays. Theorem 3 assures us (at least for reasonably large sample sizes) that the constancy behavior which characterizes the parametric S-array of a seasonal process will also be apparent in the S-array based upon the sample autocorrelation.

II. Some General Remarks

Before considering a specific example of real data, it will be helpful to consider an example which illustrates the consequences of Theorems 1 and 2. Suppose the process $\{X_t\}$ is given by

$$(1-B)^2(1-\sqrt{2}B + B^2)\phi(B)X_t = \theta(B)Z_t$$

where the zeroes of $\phi(x)$ are all outside the unit circle. Note that the equation $(1-x)^2(1-\sqrt{2}x + x^2) = 0$ has the three distinct roots 1, $\frac{\sqrt{2}}{2}(1+i)$, and $\frac{\sqrt{2}}{2}(1-i)$ which are all on the unit circle. However, since the root 1 is repeated, Theorem 1 implies that $\rho_x(m)$, the limiting autocorrelation function of $\{X_t\}$, satisfies the first order difference equation

$$y_m - y_{m-1} = 0 .$$

If $W_t = (1-B)X_t$, we can make the further observation that $\rho_w(m)$, the limiting autocorrelation function of $\{W_t\}$, satisfies the third order difference equation

$$y_m - (\sqrt{2} + 1)y_{m-1} + (\sqrt{2} + 1)y_{m-2} - y_{m-3} = 0.$$

It is also important to see that Theorem 2 implies

$$S_1(\rho_x(m)) \equiv 0 \quad , \quad S_1((-1)^m \rho_x(m)) \equiv -2,$$

and

$$S_3(\rho_w(m)) \equiv 0 \quad , \quad S_3((-1)^m \rho_w(m) \equiv -2(2+\sqrt{2}).$$

With this example in mind, we are now in a position to outline the general procedure for detecting and estimating the parameters of seasonal factors present in a process which is being observed. In the following it will be assumed that a realization $\{x_1, x_2, \ldots, x_N\}$ has been obtained from an ARMA process of the form in (2), where all of the roots of the characteristic equation $\phi(x) = 0$ lie on or outside the unit circle.

The first step in simply detecting the presence of seasonal factors (which should of course follow an initial look at a plot of the data) is an examination of the sample S-arrays (i.e., the S-arrays for $\hat{\rho}_m$ and $(-1)^m \hat{\rho}_m$). Theorem 2 implies that seasonal factors are in the process if and only if one of the following situations holds:

(i) $S_j(f_m) \equiv c$ for some $j \geq 2$

(ii) $S_1(\hat{\rho}_m) \equiv -2$

(iii) $S_1((-1)^m \hat{\rho}_m) \equiv 0$

If N is reasonably large, Theorem 3 assures us that this constancy behavior in the parametric S-array should also be evident in the sample S-array. Thus, in order to detect seasonal factors the sample S-array should be examined for the presence of a near constant column. (The interpretation of "near constant" will be elaborated on later.) Upon detecting a nonstationarity in the process, Theorem 1 suggests clearly that the next step in identifying the full order of the process is to operate on the data by the correct nonstationary operator. The reason for this is that the autocorrelation function of a nonstationary process may satisfy a difference equation of order less than p (where p is the order of $\phi(x)$); and therefore the constancy behavior expected in the p^{th} column of the sample S-array may be completely obscured by the presence of nonstationary factors. If, however, the data is transformed by the correct nonstationary operator, the resulting S-array may be examined for additional stationary or nonstationary factors. The example to be presented later should make these ideas clear.

The question of how to choose the parameters of the nonstationary (seasonal or nonseasonal nonstationary) operator now arises. In the discussion of this problem it should be understood that process factors whose zeroes are not on but only close to the unit circle will also induce a near constant column in the sample S-array. Thus, when referring to the problem of estimating the parameters of a nonstationary operator, we leave open the possibility that the appropriate operator is mathematically stationary but has zeroes which are close to the unit circle. The procedure to be suggested for choosing the parameters consists of two stages. Suppose that the k^{th} column of the sample S array is near constant, or (as sometimes occurs in practice) that the k^{th} column and one or more columns previous to the k^{th} are nearly constant. The first stage, then, in choosing the parameters of the nonstationary operator is the fitting of a k^{th} order autoregressive model to the data, estimating parameters by the Yule-Walker method (see Box and Jenkins (1976)). After fitting the model, the roots of the resulting characteristic equation should be examined. Associated with each complex conjugate pair of roots $a_j \pm ib_j$ (which may be indexed by $\lambda_j = a_j + ib_j$, $b_j > 0$) is the frequency $\omega_j = \frac{1}{2\pi} \tan^{-1}(\frac{b_j}{a_j})$ and the modulus $|\lambda_j^{-1}| = \frac{1}{\sqrt{a_j^2 + b_j^2}}$. For real roots, $\lambda_j = a_j$ and we have

$$\omega_j = \begin{cases} 0 , & a_j \geq 1 \\ \tfrac{1}{2} , & a_j \leq 1 \end{cases} \quad \text{and}$$

$$|\lambda_j^{-1}| = \frac{1}{|a_j|} .$$

The ω_j and $|\lambda_j^{-1}|$ can be used in the same way that a spectrum would be used to compare the relative contribution of each frequency to the overall variation in the process. The ω_j and $|\lambda_j^{-1}|$ have an advantage over the spectrum, however, in that the ω_j are not "smeared" as they would be in the spectrum and the $|\lambda_j^{-1}|$ can be used to determine stationarity and nonstationarity. Note that the closer $|\lambda_j^{-1}|$ is to 1 the closer the corresponding factor is to being nonstationary. The ω_j here may be thought of as the natural frequencies of the process rather than as the harmonics in a Fourier series expansion. This initial fitting of a model and the subsequent examination of roots should be used as an investigative procedure to simply identify the nature of the nonstationary (or nearly nonstationary) factors in the process being observed. The preliminary nature of the initial fit indicates why Yule-Walker estimates were suggested rather than, for example, MLEs (under the assumption $Z_t \sim N(0,\sigma^2)$) since the computing time required to calculate MLEs is usually much greater than that needed for Yule-Walker estimates.

The second stage in the parameter estimation problem involves choosing between a nonstationary and a stationary model. In part, resolving this question requires us to elaborate somewhat on the meaning of the phrase "near constant column in the sample S-array." Besides a simple examination of the k^{th} column (the "constant" column referred to previously), another method for determining the degree of constancy is to form one (or both) of the ratios

$$R_{0,k} = -\frac{\sum_{i=\ell}^{a_N} S_k(\hat{\rho}_{-k+i})}{\sum_{i=\ell-1}^{a_N-1} S_k(\hat{\rho}_{-k-i})} \quad , \quad R_{1,k} = (-1)^{k+1} \frac{\sum_{i=\ell}^{a_N} S_k((-1)^{-k+i}\hat{\rho}_{-k+i})}{\sum_{i=\ell-1}^{a_N-1} S_k((-1)^{-k-i}\hat{\rho}_{-k-i})} ,$$

where a_N is some constant which depends on the record length and $\ell \geq 2$. A rule of thumb which has proven useful in practice is to give serious consideration to a nonstationary model whenever $|R_{1,k}|^{\frac{1}{k}} \geq .95$. In order to see the rationale behind this rule of thumb recall that for a polynomial $1 + a_1x + a_2x^2 + \ldots + a_kx^k$

$$|a_k| = \prod_{i=1}^{k} |r_i^{-1}| \quad , \text{ where } r_1, r_2, \ldots, r_k \tag{3}$$

are the k zeroes of the polynomial. It can also be shown (see Woodward and Gray 1979) that

$$-\frac{S_k(\hat{\rho}_{-k+1})}{S_k(\hat{\rho}_{-k})} = (-1)^{k+1} \frac{S_k((-1)^{k+1}\hat{\rho}_{-k+1})}{S_k((-1)^k\hat{\rho}_{-k})} = \hat{\phi}_k \quad ,$$

where $\hat{\phi}_k$ is the Yule-Walker estimate for the k^{th} coefficient of an autoregressive process of order k. Because of the assumed constancy behavior in the k^{th} column, $R_{1,k}$ is obviously another (and usually better) estimate for the k^{th} coefficient, and thus if $|R_{1,k}|^{\frac{1}{k}} \geq .95$ it can be seen from (3) that most of the $|r_i^{-1}|$ for a fitted model will necessarily be close to one. (It should be remembered that we have disallowed roots <u>inside</u> the unit circle in all our discussion.) $R_{1,k}$ is thus informative for purposes of choosing between a nonstationary and a stationary model. Of course, once a k^{th} order model has been fit the $|r_j^{-1}|$ can be examined directly as has already been described. The use of $R_{1,k}$, though, is still helpful as a preliminary means of checking whether an observed "constancy" behavior is evidence of a nonstationarity in the process. Ultimately, however, the decision to fit a nonstationary model should also be based upon an understanding of the physical aspects of the time series under consideration and/or the desired nature of the forecast function.

Once the decision has been made to fit a nonstationary model, a method of estimating the parameters of the model is needed. In order to outline one method, first suppose that it has been decided to treat the factors associated with $\lambda_1, \lambda_2, \ldots, \lambda_i$ (where the λ_j's are a subset of the roots obtained in the preliminary Yule-Walker fit) as nonstationary factors. The suggested procedure for obtaining the seasonal or non-seasonal nonstationary model is to adjust these factors from

$$(1-\lambda_j^{-1}B)(1-\bar{\lambda}_j^{-1}B) = 1 - 2a_j|\lambda_j^{-1}|^2B + |\lambda_j^{-1}|^2B^2$$

to $\quad 1 - \alpha_j B + B^2 \quad$ where

$$\alpha_j = \mathrm{sgn}[\tan(2\pi\omega_j)] \frac{2}{\sqrt{\tan^2(2\pi\omega_j)+1}} .$$

Note that the frequency associated with the adjusted factor $(1-\alpha_j B+B^2)$ is still the natural frequency ω_j but that the zeroes of this factor are on the unit circle. This adjustment is analogous to what is frequently done in the special case $\omega_j = 0$ when the data is differenced. (If $\omega_j = 0$ or $\frac{1}{2}$, λ_j is real and the adjusted factors are 1-B and 1+B respectively.) Since the adjusted factor depends only upon ω_j, another possible method for obtaining the nonstationary factor is to determine ω_j from a spectral analysis. A spectral analysis alone is not sufficient to determine that a factor is nonstationary, but it may be helpful in determining ω_j precisely once it is determined that the factor associated with ω_j is nonstationary (i.e., that $|\lambda_j^{-1}| \doteq 1$).

In order to identify the full order of the model, the data should be transformed by either the operator from the initial Yule-Walker fit or the adjusted operator. Since it has been observed that the S-array values for transformed data are sometimes sensitive to slightly different transforming operators, it is not always clear which of the two operators should be used. For this reason, examining the S-arrays for both transformations is often useful for purposes of identifying the full order of the model. In practice it will sometimes happen that one transformation will induce a clear pattern of constants in some column of the S-array, whereas the other transformation will not. Since such a pattern of constants indicates that the sample autocorrelation of the transformed data nearly satisfies a difference equation, examining the S-arrays of both transformations may actually indicate which one of the original transformations provides a better fit to the data.

III. Modeling the International Airline Data

In order to illustrate the method which has just been discussed for modeling nonstationary data, we will obtain models for the well-known international airline data using this method. The airline data (see Fig. 1) was first analyzed by Box and Jenkins (1970) and is made up of 144 monthly totals of airline passengers. As Box and Jenkins have pointed out, if $\{Y_t\}$ is the airline series and $X_t = \ell n Y_t$, then the series $\{X_t\}$ is more compatible with the linearity and homoscedasticity assumptions inherent in ARMA models than is the series $\{Y_t\}$. Their model is thus for $\{X_t\}$ as will be the models in the present analysis. The Box and Jenkins model is

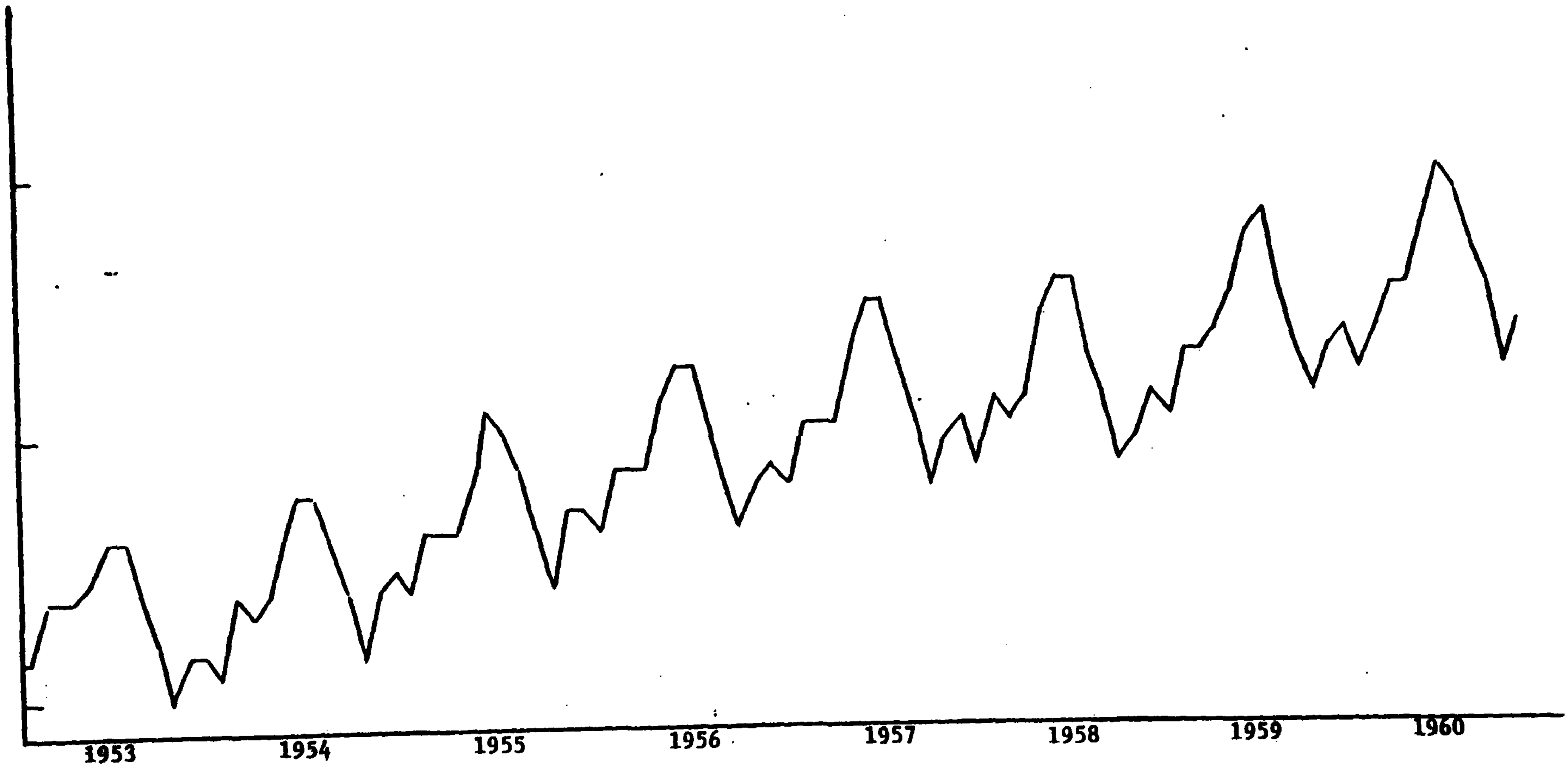


Figure 1

Airline Data

$$(1-B)(1-B^{12})X_t = (1-0.4B)(1-0.6B^{12})Z_t \qquad (4)$$

which is arrived at by a consideration of the physical aspects of the airline data. It will be shown in the present analysis how models similar to (4) may be obtained by allowing the data to "speak for itself."

Following the outline in the previous section, the first step in our analysis is an examination of the sample S-array. A portion of the S-array for $(-1)^m\hat{\rho}_m$ is shown in Table 1. Note the strong degree of constancy in the first column. With $\ell = 2$ and $a_N = 8$ we have $|R_{1,1}| = .96$, which is evidence that a nonstationary or nearly nonstationary first order factor is in the process. From our discussion earlier, an estimate for this factor is (1-.96B). The constancy behavior seen in column thirteen of the S-array may be explained after the data is transformed by (1-.96B). The S-array (using $\hat{\rho}_m$) for the transformed data is seen in Table 2. Column twelve of this array exhibits a constancy behavior which seems to be of the type which characterizes stationary autoregressive processes (see Gray, Kelley, and McIntire 1978). However, upon fitting a 12^{th} order Yule-Walker model to the transformed data and examining the roots of the resulting characteristic equation (Table 3), it is seen that the observed constancy behavior is actually indicative of an operator which is quite nearly the seasonal operator $(1-B^{12})$. From Theorem 2, if a process is of the form

$$(1-B^{12})\phi_s(B)X_t = \theta(B)Z_t$$

(where $\phi_s(B)$ is a stationary operator), then $S_{12}(\rho_m) \equiv 0$. It is now clear, then, that the values in column 12 of the sample S-array are estimates (in the sense of an average) of some value near zero, and the observed constancy behavior is thus consonant with the theory which has

	S 1	S 2	S 3	S 4		S 12	S 13	S 14
-8	-2,014	2.975	-3.102	3.816		16.835	5.391	-6.103
-7	-2,026	2.785	17.201	7.707		-44.057	5.514	-34.950
-6	-2.030	6.200	-5.517	12.865		35.093	4.747	-7.236
-5	-2.038	5,792	-93.626	6.344		-62.151	5.052	-13.236
-4	-2.052	2.658	-1.702	-2.645		15.101	4.480	1.831
-3	-2.056	4.411	10.158	9.296		-78.912	3.928	-.035
-2	-2.061	-3.979	-30.891	55.245		27.774	3.423	-.451
-1	-2.049	18.652	-42.121	-95.704	...	69.892	3.083	45.095
0	-1.954	2.182	-2.301	2.247		2.914	-1.498	1.550
1	-1.943	1.282	-3.237	13.625		35.456	-1.656	-.284
2	-1.947	3.653	-6.006	-2.586		15.087	-1.869	-.019
3	-1.950	8.614	5.411	1.904		-35.686	-2.148	.683
4	-1.964	3.041	-3.130	4.277		9.388	-2.353	3.557
5	-1.971	2.936	58.845	9.373		-26.598	-2.199	7.287
6	-1.975	7.283	-5.354	18.841		18.460	-2.507	2,901
7	-1.986	6.220	-60.891	11.057		-17.521	-2.443	25.345

Table 1

A portion of the S-array (using $(-1)^m \hat{\rho}_m$) for the international airline data previously been discussed. By considering $|R_{0,12}|^{\frac{1}{12}}$ for the transformed data the initial confusion could have been avoided. $|R_{0,12}|^{\frac{1}{12}}$ = .970 (ℓ = 2, a_N = 6), which clearly indicates that the constancy in the 12^{th} column is evidence of a seasonality in the data.

Table 3 exhibits clearly that an operator quite similar to $(1-B^{12})$ should be included in our model for the airline data. The frequencies associated with the roots of the fitted characteristic equation are quite compatible with those of $1 - x^{12} = 0$, and the root -1.0811 is the only one which might not be considered sufficiently close to the unit circle. For reasons of simplicity and parsimony, then, the models in the present analysis will include the operator $1-B^{12}$. At this point, if it is not desired to treat the factor (1-.96B) as nonstationary, a reasonable procedure would be to fit a model to the series $(1-B^{12})X_t$. The S-array for the data transformed by $(1-B^{12})$ indicates that the transformed data may be adequately modeled as a 13^{th} order, stationary autoregressive process. A reasonable initial model for the airline data would thus be

$$(1-B^{12})\phi_1(B)X_t = Z_t \quad , \tag{5}$$

where

$$\begin{aligned}\hat{\phi}_1(B) = 1 &- .535B - .272B^2 + .057B^3 - .018B^4 - .087B^5 \\ &- .041B^6 + .076B^7 - .038B^8 - .158B^9 + .136B^{10} \\ &+ .155B^{11} + .287B^{12} - .295B^{13} \ .\end{aligned}$$

An estimate of the variance of Z_t for this model is $\hat{\sigma}_z^2 = .00127$. If the original data is transformed by $(1-B)(1-B^{12})$, a 12^{th} order, stationary autoregressive process is a satisfactory model for the transformed data. Another contending model would thus be

$$(1-B)(1-B^{12})\phi_2(B)X_t = Z_t \tag{6}$$

where

	S 1	S 2	S 3	S 4	S 5	S 6	S 7	S 8	S 9	S 10	S 11	S 12
-6	-2.397	-11.704	-1.374	-.308	.020	1.738	-1.868	3.662	-4.377	11.502	-7.059	-.496
-5	3.772	4.270	1.712	.184	-2.090	1.993	-1.621	3.555	28.772	9.605	22.143	-.750
-4	-.565	.441	-1.327	2.934	-3.078	3.169	-2.992	4.346	-3.191	5.679	-4.630	-.606
-3	-.237	4.060	.093	4.702	-3.959	-.356	17.532	3.779	-13.592	6.229	29.656	-.999
-2	-3.146	1.948	-4.922	4.710	-11.772	-17.885	20.560	4.833	-4.351	12.642	-7.176	-.679
-1	3.709	6.010	-12.421	4.408	27.276	34.430	-8.763	4.941	-60.950	9.688	15.388	-.817
0	-.788	.906	-.978	1.257	-1.201	1.245	-1.451	2.054	-2.125	2.722	-2.313	.604
1	-1.466	.505	2.284	1.395	-2.231	-4.249	1.588	1.953	14.403	3.769	10.767	.497
2	.311	-1.380	.063	1.338	7.045	-.207	-10.261	1.593	-2.472	8.201	-5.750	.703
3	1.300	-2.809	-1.364	1.323	2.506	10.053	-8.985	1.785	9.695	6.958	20.154	.420
4	-.790	.911	-.724	-.117	-.289	-.531	-.556	3.508	-3.048	4.564	-3.633	.506
5	-1.716	2.532	1.201	.263	-.026	-.127	-1.909	3.424	-18.237	5.208	14.444	.331

Table 2

S-array (using $\hat{\rho}_m$) for data transformed by (1-.96B)

Table 3

Roots (λ_i) of characteristic equation from model fitted to transformed data

λ_i	frequency (λ_i)	$\lvert\lambda_i^{-1}\rvert$
1.0607	0	.9427
.8747 ± (.5039)i	.0832	.9906
.5012 ± (.8757)i	.1673	.9911
.0165 ± (1.0285)i	.2474	.9721
-.5091 ± (.8763)i	.3338	.9867
-.8912 ± (.5025)i	.4183	.9774
-1.0811	.5000	.9250

Roots (r_i) of $1 - x^{12} = 0$

r_i	frequency (r_i)	$\lvert r_i^{-1}\rvert$
1	0	1
.8660 ± (.50)i	.0833	1
.50 ± (.8660)i	.1667	1
±i	.2500	1
-.50 ± (.8660)i	.3333	1
-.8660 ± (.50)i	.4167	1
-1	.5000	1

$$\hat{\phi}_2(B) = 1 + .358B + .054B^2 + .151B^3 + .110B^4$$
$$-.047B^5 - .089B^6 + .015B^7 - .031B^8 - .164B^9$$
$$-.036B^{10} + .081B^{11} + .339B^{12} \text{ and } \hat{\sigma}_Z^{\ 2} = .00136 \ .$$

For illustrative purposes $\hat{\phi}_1(B)$ and $\hat{\phi}_2(B)$ were fit by the Yule-Walker method, although MLEs may be desirable for a more refined model. Approximate standard errors for the coefficients in these operators are given in Table 4. (The standard errors were found under the assumption that $Z_t \sim N(o,\sigma^2)$ by using the approach of Box and Jenkins (1976).)

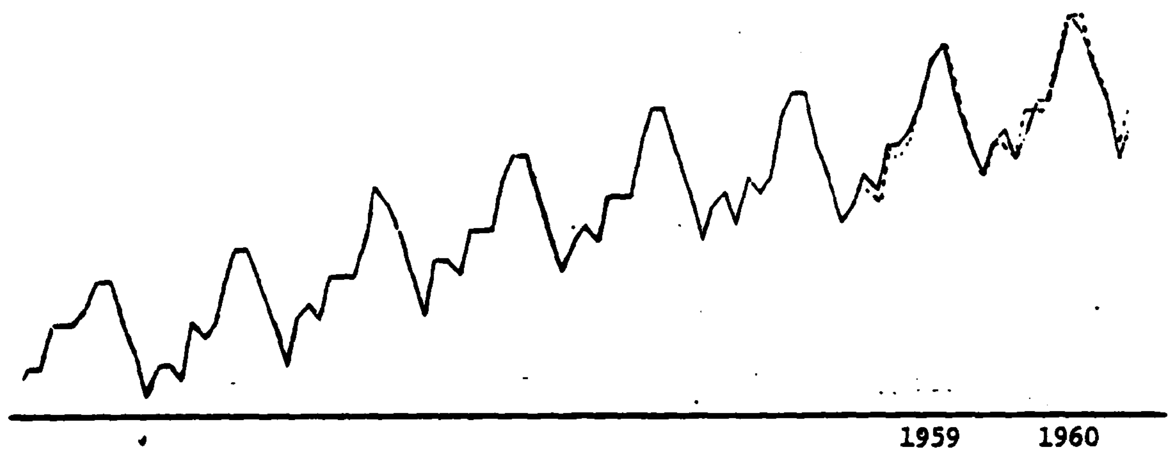
It is clear that many other reasonable models could be obtained by using different operators of the form $(1-\phi B)\,\phi_{12}(B)$ where $\phi_{12}(B)$ is some 12th order operator similar to $1-B^{12}$. Models (5) and (6) certainly seem adequate, however, and they illustrate well the method for modeling seasonal data through S-arrays. Note that model (6) has two roots of unity and will thus have a forecast function possessing a linear trend; whereas the forecast function for model (5) will not contain the linear trend for forecasts at long lead times since, instead of two units roots, this model has one unit root, one root equal to 1.345, and an additional low frequency component. In general, this kind of a difference between the forecast functions of stationary and nonstationary models can provide a means of choosing between competing models.

A comparison of forecasts obtained from models (4) and (5) may be seen in Fig. 2-4. The forecasts are made from origins 24 months, 36 months, and 48 months prior to the end of the data set. SSE (i.e. $\sum_{i=1}^{k}(x_{t_o+i}-\hat{x}_{t_o+i})^2$) for model (4), the Box-Jenkins model, is seen to be more erratic than that of model (5) in the sense that it does not increase monotonically as the maximum number of steps ahead to be forecast increases. The reason for this appears to be that the trend component of the forecast function for model

Table 4

	Coefficient	Estimate	Standard Error
	$\phi_{1,1}$	.535	.083
	$\phi_{1,2}$	.272	.092
	$\phi_{1,3}$	-.057	.094
	$\phi_{1,4}$	.018	.093
	$\phi_{1,5}$	.087	.092
	$\phi_{1,6}$	.041	.093
Model (5)	$\phi_{1,7}$	-.076	.092
	$\phi_{1,8}$	.038	.093
	$\phi_{1,9}$	.158	.092
	$\phi_{1,10}$	-.136	.093
	$\phi_{1,11}$	-.155	.094
	$\phi_{1,12}$	-.287	.092
	$\phi_{1,13}$	.295	.083
	$\phi_{2,1}$	-.358	.082
	$\phi_{2,2}$	-.054	.088
	$\phi_{2,3}$	-.151	.088
	$\phi_{2,4}$	-.110	.088
	$\phi_{2,5}$	.047	.088
Model (6)	$\phi_{2,6}$	.089	.088
	$\phi_{2,7}$	-.015	.088
	$\phi_{2,8}$	.031	.088
	$\phi_{2,9}$	.164	.088
	$\phi_{2,10}$	.036	.088
	$\phi_{2,11}$	-.081	.088
	$\phi_{2,12}$	-.339	.082

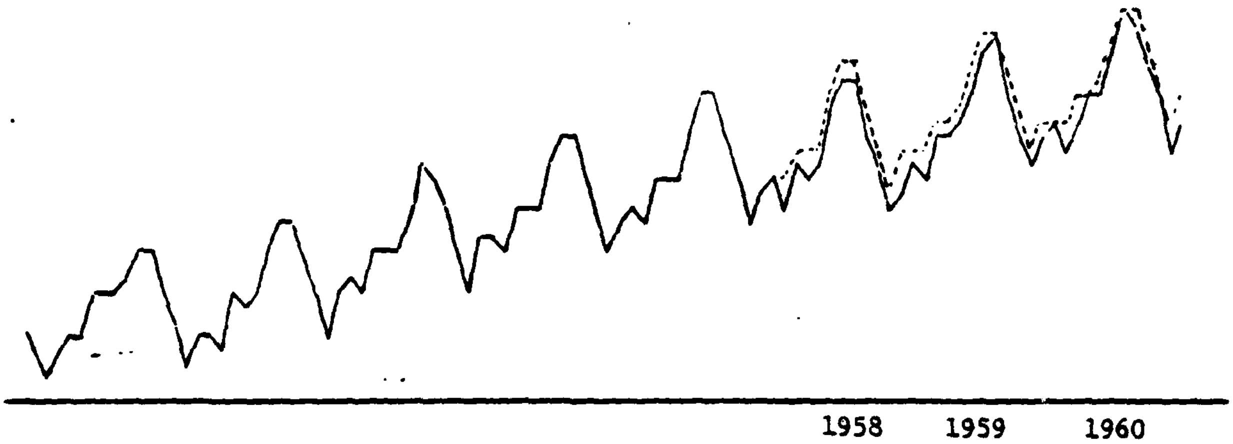
Model (4) SSE = .6888



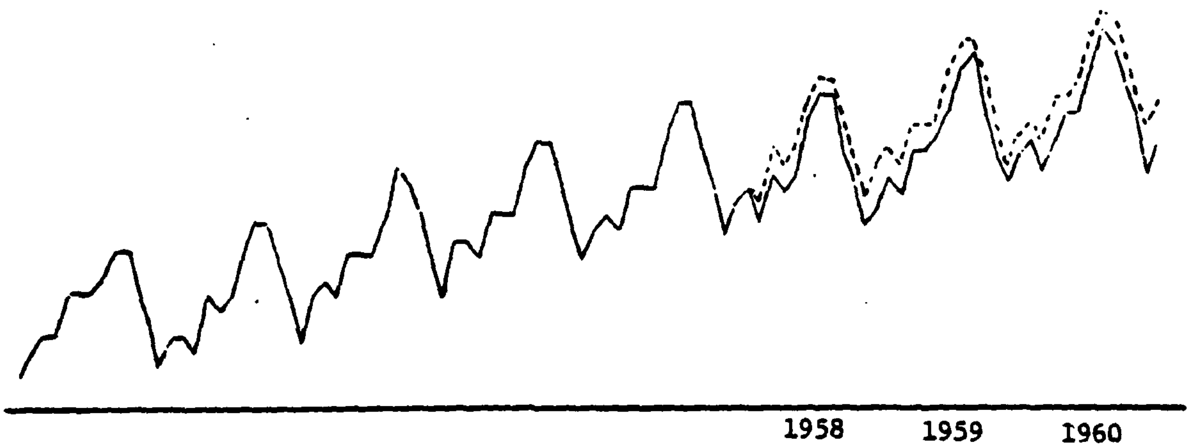
Model (5) SSE = .0203

Figure 2

Airline Data 24 Month Forecasts



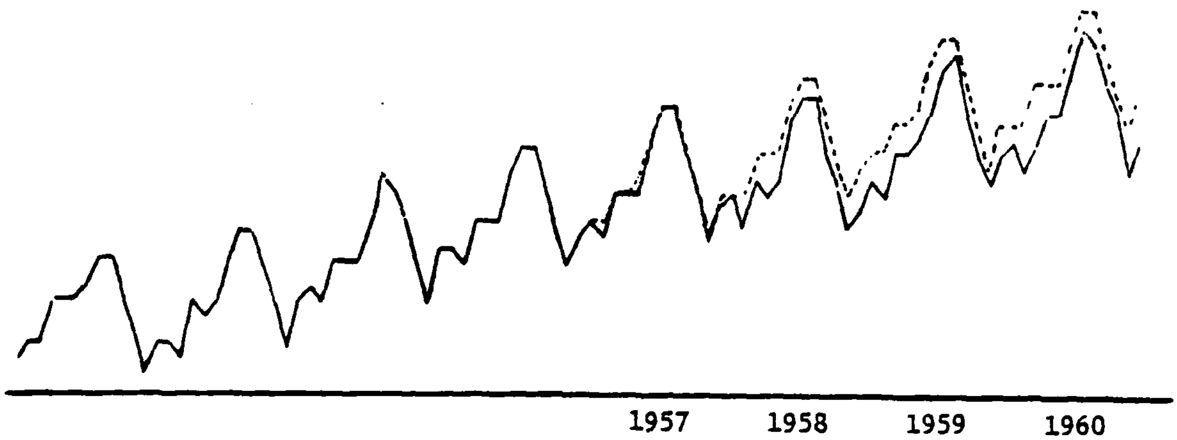
Model (4) SSE = .1656



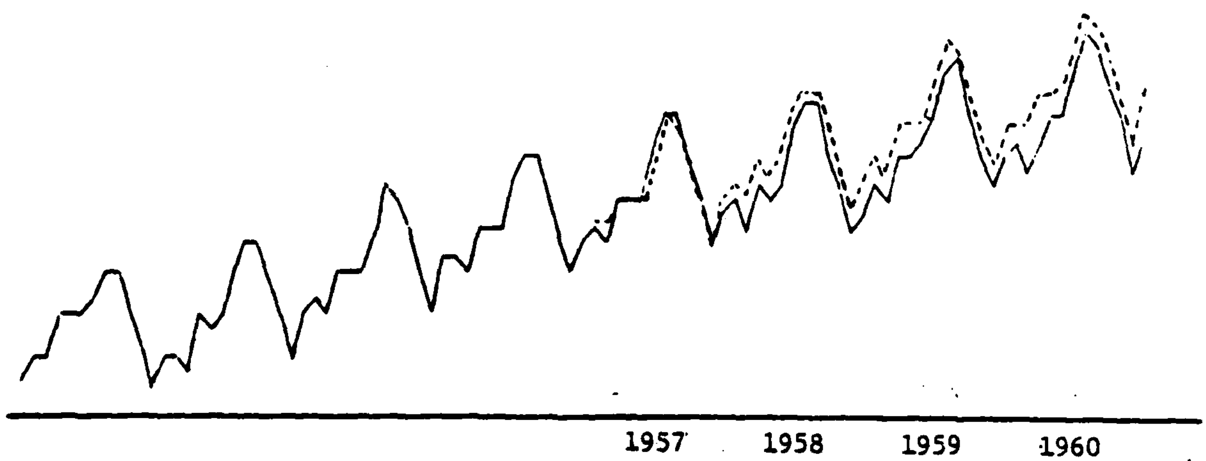
Model (5) SSE = .3066

Figure 3

Airline Data 36 Month Forecasts



Model (4) SSE = .4627



Model (5) SSE = .3335

Figure 4

Airline Data 48 Month Forecasts

(4) is determined by only 13 values previous to the forecast origin; whereas the corresponding component for model (5) is determined by 25 preceding values. Since a close inspection of the data reveals evidence of a 24 month period, it is understandable why a forecast function using only 13 months prior to the forecast origin might perform eratically. When one considers that model (5) was fit through a purely data analytic approach it is not surprising that its forecast function contains the low frequency component not found in model (4).

For an interesting and more complete discussion of forecasts obtained from (4), (5), and a model proposed by Parzen see Gray and Woodward (1980).

IV. Conclusion

A method of modeling nonstationary ARMA processes (with special emphasis on seasonal processes) has been examined in this paper. A brief outline of this method is as follows:

(i) Detect the presence of nonstationary factors by examining the sample S-array.

(ii) Determine the nature of the detected nonstationary factors by fitting an appropriate Yule-Walker model to the data.

(iii) Transform the data by the operator obtained from the Yule-Walker fit or an adjusted operator and examine the S-array of the transformed data for the presence of additional stationary or nonstationary factors.

(iv) After identifying the full order of the model, decide which factors are to be treated as nonstationary. Transform the data by the nonstationary factors and fit a model of appropriate order to the transformed data.

Using this methodology, it has been shown how models for the international airline data may be obtained through a data analytic technique rather than by simply considering the physical aspects of the data.

REFERENCES

Box, G. E. P. and Jenkins, G. M. (1976), Time Series Analysis; Forecasting and Control (Rev. ed.), San Francisco: Holden-Day.

Findley, D. F. (1978), "Limiting Autocorrelations and their Uses in the Identification of Nonstationary ARMA Models," Bulletin of the Institute of Mathematical Statistics 7, 5, 293-294.

Gray, H. L., Kelley, G. D., and McIntire, D. M. (1978), "A New Approach to ARMA Modeling," Communications in Statistics B7, 1-78.

Morton, Michael (1980), "Seasonal ARMA Models," Southern Methodist University, Department of Statistics Technical Report No. 141, Dallas, Texas.

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

Woodward, W. A. and Gray, H. L. (1979), "On the Relationship between the R- and S-Arrays and the Box-Jenkins Method of ARMA Model Identification," Southern Methodist University, Department of Statistics Technical Report No. 134, Dallas, Texas.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 140	2. GOVT ACCESSION NO. AD-A092 515	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Modeling Seasonal ARMA Processes		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER 140
7. AUTHOR(s) Jeffrey D. Hart H. L. Gray		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0439
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Dallas, Texas 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042 280
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, Va. 22217		12. REPORT DATE October 1980
		13. NUMBER OF PAGES 26
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purposes of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Gray, Kelley, and McIntire (1978) have introduced a method, based on arrays of numbers called R- and S- arrays, for identifying p and q in an ARMA (p,q) process. In addition, they have illustrated how the same method is useful in detecting nonstationary factors in an observed process, and in suggesting an appropriate transformation to stationarity. In the present paper special attention is given to the problem of modeling seasonal ARMA processes using the S-array method. A general definition is given for a seasonal process,

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (con't)

and the procedure for identifying and modeling such processes is discussed in detail. Additionally, an interesting theorem characterizing the S-arrays (based upon the sample autocorrelation) of seasonal processes is stated and a proof indicated. Finally, a data set (the international airline data) which exhibits the properties of a seasonal process is analyzed using the method discussed, and two models for the data are proposed.