AN AUTOMATED ANALYSIS OF TIME SERIES

by

KEVIN JAMES HELLMAN, B.S.

A THESIS

IN

STATISTICS

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

MASTER OF SCIENCE

Approved

Accepted

May, 1980

ACKNOWLEDGEMENTS

I would like to thank Dr. Thomas L. Boullion for his guidance of this thesis, and Dr. Benjamin S. Duran for his encouragement and support. Special thanks go to Dr. John Jensen for his help with the computer details, and to Dunree Norris for the typing.

TABLE OF CONTENTS

I. INTRODUCTION

Due to the stochastic nature of sales, figures, stock prices, market values, and other commonly reviewed data, one would have to agree that sales managers, systems analysts, researchers, and many others would find time series analysis an invaluable aid in decision making and forecasting. However, the complexity that is inherent in the techniques used to identify, estimate, and forecast a time series makes anything other than the most elementary methods beyond the ability of those without a reasonable background in mathematics and statistics. Therefore, it is the objective of this thesis to present a practical computer package which is to give results of the analysis of a time series which are understandable and interpretable by any user. Such a package, then, will solve such problems as:

i) the detection and removal of any non-stationary behavior

ii) the determination of the order of a parsimonious ARMA model to adequately represent the series

iii) the computation of least squares estimates for the parameters of the model.

iv) the diagnostic checking of the adequacy of the model

v) the computation of least squares forecasts and confidence limits for these forecasts.

The methods of dealing with non-stationary series will be presented in the second part of the thesis. All other procedures of the package will make use of methods proposed by others, and will be explained in the first part of the thesis.

II. DESCRIPTIONS OF METHODS

A. THE ARMA MODEL

The package is designed to fit an ARMA (p,q) model to the stationary time series. That is, the series is modelled as

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p}$$

$$+a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

which can also be written as

$$\phi(B) Z_t = \theta(B) a_t$$

where

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

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

and where B denotes an operator such that

$$BZ_t = Z_{t-1}$$

with Z_t denoting the observed value of the series at time t, and with $\{a_t\}$ being a sequence of uncorrelated "shocks" which have mean zero and a finite variance, and which are assumed to be at least approximately normally distributed. ϕ_i, i=1,...,p and θ_j, j=1,...,q denote the parameters which are to be estimated.

Since a parsimonious representation is desired, and because some limits are necessary in any computer package, p and q of the ARMA model are each arbitrarily limited to be at most three. This is not thought to be too serious a limitation since most series encountered can be parsimoniously represented by a third order model.

B. THE R AND S ARRAYS

Gray, Kelly, and McIntire [4] introduced a method for determining the order of an ARMA process which was based on the R and S arrays. These arrays are computed by defining r_m to be the estimate of the autocorrelation of the series at lag m, then letting

$$f_m = (-1)^m r_m,$$

(2.1)

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

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

Then, let

$$R_n(f_m) = \frac{H_n(f_m)}{H_{n+1}(1;f_m)} ,$$

(2.2)

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

Recursive formulae for the calculation of $R_n(f_m)$ and $S_n(f_m)$ are given in [4], p.3. For the remainder of the thesis, the simpler notations $R_n(m)$ and $S_n(m)$ will be used in place of $R_n(f_m)$ and $S_n(f_m)$, respectively.

C. NON-STATIONARY OPERATORS

As defined in [4], p.23, a series is non-stationary whenever one or more roots of

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

lies on the unit circle. Of course, if a root of $\phi(B)$ is "almost" on the unit circle, the series behaves like a non-stationary series for modelling purposes. Now, the S array has the property that a root or roots of $\phi(B)$ on the unit circle causes the column of the S array which corresponds to the number of "unit" roots to become constant. Due to the fact that estimated autocorrelations must be used to calculate the S-array, the "constant" behavior of the column is actually almost constant behavior, which is also noticeable for roots near the unit circle. The method used to determine if a column of the S array is constant enough to assume unit roots is proposed in the second part of the thesis. Once a unit root is assumed, the behavior of the S array column indicates the particular non-stationary operator present. Thus,

$$S_1(m) \equiv -2 \quad m = 0, \pm 1, \ldots$$

is indicative of the operator (1-B).

$$S_1(m) = 0 \quad m = 0, \pm 1, \ldots$$

occurs if an operator (1+B) is present.

$$S_2(m) = C \neq 0 \quad m = 0, \pm 1, \ldots$$

indicates an operator $(1-aB+B^2)$, where a = C-2.

$$S_2(m) = 0 \quad m = 0, \pm 1, \ldots$$

indicates the presence of the operator $(1-B^2)$. Obviously, many more possibilities exist, but it is assumed that only non-stationary operators of order two or less will be present in any series encountered in practice.

D. IDENTIFICATION OF ORDER

Another property of the S array is that, for stationary series, the column which corresponds to p, the order of the autoregressive operator, has two distinct constant stretches. As before, the behavior is actually that of being almost constant. Also, the number of non-constant values, that is, entries obviously not equal to either of the constant values, is equal to 2q, where q is the order of the moving average operator. In order for the package to decide on the values of p and q, the D-statistic, defined in [4], p. 20, is used. It is there noted that for ARMA processes, the D-statistic is superior to other automated methods of identification, although not necessarily better in the case of strictly AR processes. Also noted [4] is that the D-statistic is not normally of any use with small sample sizes. Therefore, an arbitrary lower limit of one hundred observations will be required for the package, so that the D-statistic can be expected to perform satisfactorly.

E. INITIAL ESTIMATES OF THE PARAMETERS

Once the order of the model has been determined, initial estimates of the parameters are required in order to start off the iterative least-squares procedure. The method used to find initial estimates is explained in detail in [3] p. 359. In this method the stationary ARMA model

$$\phi(B) Z_t = \theta(B) a_t$$

is rewritten in a strictly AR form

$$\pi(B) Z_t = a_t$$

where

$$\pi(B) = 1 - \sum_{i=1}^{\infty} \pi_i B^i .$$

Truncation of the infinite sume after $k = p + q$ terms makes it possible to define the $k \times k$ system

$$A_n \pi = V_n , \tag{2.3}$$

where

$$\pi = (\pi_1, \pi_2, \ldots, \pi_k)',$$

$$A_n = \frac{1}{n-k} \sum_{t=k+1}^{n} X_t' X_t ,$$

$$V_n = \frac{1}{n-k} \sum_{t=k+1}^{n} X_t' Z_t ,$$

and X_t is defined by

$$X_t = (Z_{t-1}, \ldots, Z_{t-k+1}).$$

The system (2.3) is then solved for π. Next, the $k \times k$ system

$$\pi_1 = \theta_1 - \phi_1$$

$$\pi_2 = \theta_1 \pi_1 + \theta_2 - \phi_2$$

$$\pi_j = \sum_{i=1}^{\min(j,q)} \theta_i \pi_{j-i} - \phi_j, \; j \leq p$$

$$\pi_j = \sum_{i=1}^{\min(j,q)} \theta_i, \pi_{j-i}, \; j > p$$

is solved for $\theta_1, \ldots, \theta_q, \phi_1, \ldots, \phi_p$, which are then used as initial estimates.

F. LEAST SQUARES ESTIMATES

A method described in [6] and credited to Hartley is used to calculate the least squares estimates of the parameters. The residuals needed in the iterations are found by backforecasting to obtain the "shocks" prior to t=0, and then proceeding forward to find the residuals. (See [1], p. 216) Now, a correction vector d is found by defining

$$\beta = (\theta_1,\ldots,\theta_2,\phi_1,\ldots,\phi_p),$$

and approximations to the derivitives needed are given by

$$X_{i,t} \doteq [(a_t|\beta_1,\ldots,\beta_i,\ldots,\beta_k) - (a_t|\beta_1,\ldots,\beta_i+\delta_i,\ldots,\beta_k)]/\delta_i$$

where $k = p + q$, a_t is the residual value for Z_t, and δ_i is arbitrarily set at .005. Then, the k × k matrix A is formed with

$$A_{ij} = \sum_{t=Q'}^{n} X_{i,t}\, X_{j,t}\ ,$$

and the k-dimensional vector g is formed with

$$g_i = \sum_{t=Q'}^{n} X_{i,t}\, a_t\ .$$

Here, Q' denotes the negative origin from which the residual calculation was begun. Now, the system

$$Ad = g$$

is solved for d. Next, a scaling vector V_m is calculated according to

$$V_m = 1/2 + 1/4\ \frac{SS(0) - SS(1)}{SS(0)-255(1/2)+55(1)}\ ,$$

where

$$SS(i) = \sum_{t=Q'}^{n} (a_t|\beta + id)^2\ .$$

Let

$$h = V_m d$$

and, letting β_0 denote the current values of the parameters, define

$$\beta = \beta_0 + h.$$

The process is repeated until convergence has occured, where three decimal places is considered adequate in the package.

G. DIAGNOSTIC CHECK

In order to affirm the adequacy of the fitted model, the "portmanteau", or general, test for goodness of fit, suggested in [1], pp. 290-293, is applied to the residuals. The estimated autocorrelations of the residuals obtained from the least squares estimated model are calculated and denoted $r_k(\hat{a})$, where k is the lag, k = 1,2,...,K. Then, define the statistic Q by

$$Q = n \sum_{i=1}^{K} r_k(\hat{a}).$$

Now, Q is distributed approximately chi-square with K-p-q-m degrees of freedom, where m = 0 if the overall constant term of the series is assumed to be zero, and m = 1 if the overall constant term is estimated by

$$\theta_0 = \overline{Z}(1 - \sum_{i=1}^{p} \phi_i).$$

It is noted in [1], p. 33 that the lag of the estimated autocorrelation function should not exceed N/4, so that the previous limitation of one hundred data points allows the number of lags to be K = 25. Due to the third order limitations imposed on p and q, there are then seven possible degrees of freedom to consider. The χ^2 percentage points for 95% confidence are therefore included in the package, and a test of goodness of fit is performed.

H. FORECASTING

The forecasting of future values of the series is accomplished via the "difference equation" approach explained in [1], pp. 129-132. That is, for lead time ℓ, the predicted value of $Z_{t+\ell}$ is given by

$$\hat{Z}_t(\ell) = \psi_1[Z_{t+\ell-1}]+\cdots+\psi_{p+d}[Z_{t+\ell-p-d}]$$

$$- \theta_1[a_{t+\ell-1}]-\cdots-\theta_q[a_{t+\ell-q}] + [a_{t+\ell}]$$

where

$$\psi(B) = \gamma(B)\phi(B),$$

with $\gamma(B)$ representing the operator of order d obtained by multiplication of all non-stationary operators, and where

$$[Z_{t-j}] = Z_{t-j} \qquad j = 0,1,2,\ldots,$$

$$[Z_{t+j}] = \hat{Z}_t(j) \qquad j = 1,2,\ldots,$$

$$[a_{t-j}] = Z_{t-j} - \hat{Z}_{t-j-1}(1) \qquad j = 0,1,2,\ldots,$$

$$[a_{t+j}] = 0 \qquad j = 1,2,\ldots .$$

The maximum lead time ℓ is arbitrarily limited to twentyfour, and the forecasts are presented with 95% confidence limits, which are calculated as in [1], pp. 132-134.

I. EXAMPLES OF PRINTOUTS

Three examples are included on the next three pages, each series being found in [1] and termed Series A, Series C, and Series E. In each example, the printing of the R and S arrays has been suppressed.

Series A, Box & Jenkins

Operate on Series with (1 - B)

and Repeat Analysis

D Statistic

Order of AR	Order of MA 0	1	2	3
0	0.2729E 01	0.1820E 04	0.5697E-01	0.1873E 03
1	0.2807E-02	0.3108E-03	0.1460E 00	0.5629E-03
2	0.6581E-02	0.7431E-01	0.7382E-03	0.2620E-02
3	0.6531E-02	0.6180E-04	0.3496E-03	0.1150E-02

Order of autoregression selected 0
Order of moving average selected 1

Initial estimates for the 1 MA parameters

-0.414

Least squares estimates of 1 MA parameters

0.707

Estimate of white noise variance

0.101

Model is adequate at the 95 percent confidence level

Lead time	1	2	3	4	5	6
.95 Limit(+/-)	0.62	0.65	0.68	0.70	0.72	0.75
Forecast	17.54	17.54	17.54	17.54	17.54	17.54

Lead time	7	8	9	10	11	12
.95 Limit(+/-)	0.77	0.79	0.81	0.83	0.85	0.87
Forecast	17.54	17.54	17.54	17.54	17.54	17.54

Lead time	13	14	15	16	17	18
.95 Limit(+/-)	0.89	0.91	0.93	0.94	0.96	0.98
Forecast	17.54	17.54	17.54	17.54	17.54	17.54

Series C, Box & Jenkins

Operate on Series with (1 - B)

and Repeat Analysis

D Statistic

Order of AR	Order of MA 0	1	2	3
0	0.6749E 00	0.9171E 00	0.1199E 01	0.1385E 01
1	0.2055E 05	0.4281E 00	0.5555E-02	0.1648E 01
2	0.3916E-03	0.2363E-02	0.1404E-04	0.2840E-04
3	0.1365E-04	0.4020E-04	0.4349E 01	0.1600E 00

Order of autoregression selected 1
Order of moving average selected 0

Initial estimates for the 1 AR parameters

0.805

Least squares estimates of 1 AR parameters

0.821

Estimate of white noise variance

0.018

Model is adequate at the 95 percent confidence level

Lead time	1	2	3	4	5	6
.95 Limit(+/-)	0.27	0.55	0.86	1.19	1.51	1.83
Forecast	18.64	18.50	18.39	18.30	18.22	18.16

Lead time	7	8	9	10	11	12
.95 Limit(+/-)	2.14	2.45	2.74	3.03	3.30	3.57
Forecast	18.11	18.08	18.04	18.01	17.99	17.97

Lead time	13	14	15	16	17	18
.95 Limit(+/-)	3.82	4.07	4.31	4.54	4.76	4.97
Forecast	17.95	17.94	17.93	17.92	17.91	17.91

Series E, Box & Jenkins

D Statistic

Order of AR	Order of MA 0	1	2	3
0	0.9630E 00	0.4029E 01	0.3040E 02	0.9483E-01
1	0.1580E-01	0.1187E-01	0.2641E-02	0.6748E 00
2	0.1707E 03	0.2325E 00	0.3798E 02	0.6926E-02
3	0.2475E-02	0.8721E-01	0.1879E-04	0.4887E 00

Order of autoregression selected 2
Order of moving average selected 0

Initial estimates for the 2 AR parameters

1.405 -0.711

Overall constant estimate

14.306

Least squares estimates of 2 AR parameters

1.422 -0.727

Estimate of white noise variance

344.666

Model is adequate at the 95 percent confidence level

Lead time	1	2	3	4	5	6
.95 Limit(+/-)	36.39	63.26	78.89	84.20	84.54	85.20
Forecast	78.34	57.61	24.98	-6.35	-27.20	-34.06

Lead time	7	8	9	10	11	12
.95 Limit(+/-)	87.66	90.28	91.64	91.88	91.90	92.21
Forecast	-28.66	-16.01	-1.92	8.90	14.50	13.52

Lead time	13	14	15	16	17	18
.95 Limit(+/-)	92.69	93.03	93.13	93.13	93.16	93.24
Forecast	9.01	2.98	-2.31	-5.45	-6.07	-4.67

III. THE DETECTION OF NON-STATIONARY OPERATORS

In order to determine if a series contains a non-stationary operator, the package must be able to decide if a column of the S array is "constant". Since the columns will never actually be constant, it is desirable to be able to test

$$H_0: \sigma^2 \leq \sigma_0^2$$

where σ_0^2 is the variance of the computed column, given that a non-staionary operator is present. If the package is expected to detect first and second order non-stationary operators, three cases must be considered.

A. ONE ROOT ON THE UNIT CIRCLE

Suppose the process

$$\phi(B)Z_t = \theta(B)a_t$$

has one root of its characteristic equation approaching the unit circle. That is,

$$\phi(B) = 0,$$

which may also be written as

$$\prod_{i=1}^{p} (1 - G_iB) = 0$$

has one root, say without loss of generality G_1, approaching the unit circle. Assume $G_1 \to 1$. Then, by Theorem 8 of [4], p. 25, the autocorrelations p_k satisfy a first order difference equation with a root of unity. That is,

$$(1-B)p_k = 0 \qquad k = 0, \pm 1, \pm 2, \ldots$$

which implies

$$p_k = p_{k-1} \qquad k = 0, \pm 1, \pm 2, \ldots \quad .$$

However, because $p_0 = 1$ by definition,

$$p_k \rightarrow 1 \qquad k = 0, \pm 1, \pm 2, \ldots \quad .$$

Now, the stationary series

$$\phi(B) Z_t = \theta(B) a_t$$

can be written

$$Z_t = \phi^{-1}(B)\theta(B) a_t$$

or

$$Z_t = \psi(B) a_t,$$

that is,

$$Z_t = \sum_{j=-\infty}^{\infty} \psi_j a_{t-j} \quad .$$

Hence, the conditions of Theorem 6.2.3. of [3], p. 240, are satisfied, so that

$$\begin{aligned} \mathrm{Cov}(r_k, r_h) = 1/N \sum_{i=-\infty}^{\infty} & [p_i p_{i-h+k} + p_{i+k} p_{i-h} \\ & - 2p_k p_i p_{i-h} - 2p_h p_i p_{i-k} \\ & + 2p_h p_k p_i^2] + 0(n^{-2}), \end{aligned} \tag{3.1}$$

and

$$E(r_k) = \frac{n-k}{n} p_k - \gamma_0^{-1}[1 - p_k]\mathrm{Var}\ (\bar{Z})$$

(3.2) $$+ \gamma_0^{-2}[p_k \mathrm{Var}(\tilde{\gamma}_0) - \mathrm{Cov}(C_0, C_k)]$$

$$+ 0(n^{-2}),$$

where γ and C correspond to the notation used for the autocovariance and estimated autocovariance of the series as in [1], and where $\tilde{\gamma}_0$ is that used in [3]. Now, if n is large, it is clear that

$$p_k \to 1 \qquad k = 0, \pm 1, \pm 2, \ldots .$$

implies

(3.3)
$$\mathrm{Cov}(r_{k+1}, r_k) \to 0,$$
$$\mathrm{Var}(r_k) \to 0,$$
$$E(r_k) \to \frac{n-k}{n} .$$

Since

$$S_1(k) = -\frac{r_{k+1}}{r_k} - 1$$

by definition,

$$E[S_1(k)] = -E[\frac{r_{k+1}}{r_k}] - 1 .$$

Using the expansion given in [5], p. 181,

$$E[\frac{r_{k+1}}{r_k}] \approx \frac{E(r_{k+1})}{E(r_k)} - \frac{1}{E^2(r_k)} \mathrm{Cov}(r_k, r_{k+1})$$

$$+ \frac{E(r_{k+1})}{E^3(r_k)} \mathrm{Var}(r_k),$$

and letting $p_k \to 1$, $k = 0, \pm 1, \pm 2, \ldots$, (3.3) implies

$$E[\frac{r_{k+1}}{r_k}] \approx \frac{E(r_{k+1})}{E(r_k)}$$

and

$$E[S_1(k)] \approx -2 + \frac{1}{n-k} .$$

Now,

$$\begin{aligned} Var[S_1(k)] &= E[S_1(k) + 2 - \frac{1}{n-k}]^2 \\ &= E[-\frac{r_{k+1}}{r_k} + 1 - \frac{1}{n-k}]^2 \\ &= E[\frac{r_{k+1}}{r_k}]^2 - (2 - \frac{2}{n-k}) E[\frac{r_{k+1}}{r_k}] + 1 - \frac{2}{n-k} + \frac{1}{(n-k)^2} \end{aligned}$$

Assuming that $p_k \to 1$, $k = 0, \pm 1, \pm 2, \ldots$ implies

$$E[\frac{r_{k+1}}{r_k}]^2 \to 1,$$

then

$$\begin{aligned} Var[S_1(k)] &\approx 1 - (2 - \frac{2}{n-k})(1 - \frac{1}{n-k}) + 1 - \frac{2}{n-k} + \frac{1}{(n-k)^2} \\ &\approx \frac{2}{n-k} - \frac{1}{(n-k)^2} . \end{aligned}$$

Therefore, define

$$\begin{aligned} X_k &= \frac{S_1(k) + 2 - \frac{1}{n-k}}{\sqrt{\frac{2}{n-k} - \frac{1}{(n-k)^2}}} \\ &= \frac{(n-k)[S_1(k)+2] - 1}{\sqrt{2(n-k) - 1}} \end{aligned}$$

for $k = 0, \pm 1, \pm 2, \ldots$. It then follows that the X_k constitute a sample from an approximate standard normal distribution. Hence, the decision that the S_1 column is constant is equivalent to testing the hypothesis

$$H_0: \sigma_x^2 \leq 1.$$

Because moving average operators affect the behavior of the "middle" entries of the S array columns, the set I_m is defined by

$$I_m = \{-m, \ldots, -4, 3, \ldots, m-1\}.$$

In this way, the effects of up to a third order moving average operator are not involved. Now, suppose I_m contain k elements. Define

$$K_1 = \frac{k-1}{k} \sum_{i \varepsilon I_m} (x_i - \bar{x})^2 .$$

Then K_1 is distributed approximately as a chi-square with k-1 degrees of freedom. Therefore, the rule for accepting or rejecting H_0 can be stated as: reject H_0 at the $(1-\alpha)\%$ confidence level if $K_1 > X^2_{k-1,(1-\alpha)}$. Equivalently, assume S_1 is constant if H_0 is not rejected. As an example, consider the series in [1] referred to as Series C. The first columns of the S array for the series and its first difference, and the corresponding X's are:

Series C

i	$S_1(i)$	x_i
-7	-2.080	-.911
-6	-2.073	-.834
-5	-2.064	-.689
-4	-2.056	-.601
3	-1.946	.523
4	-1.939	.596
5	-1.931	.679
6	-1.925	.740

First Difference

i	$S_1(i)$	x_i
-7	-2.217	-2.386
-6	-2.192	-2.112
-5	-2.163	-1.797
-4	-2.190	-2.082
3	-1.840	1.640
4	-1.859	1.484
5	-1.838	1.701
6	-1.821	1.875

Since K_1 for Series C is 3.5, K_1 for the differenced series is 25.1, and $\chi^2_{7,.95} = 14.1$, Series C should clearly be differenced once. This is in complete agreement with both [1] and [4].

The case where $G_1 \to -1$ is similar, because this would imply $p_k \to (-1)^k$, $k = 0, \pm1, \pm2, \ldots$. However, (3.1) still implies that

$$\text{Cov}(r_{k+1}, r_k) \to 0$$

and

$$\text{Var}(r_k) \to 0.$$

Now, from (3.2) it is clear that, for even k, $p_k \to (-1)^k = 1$. Hence, if k is even, $E(r_k)$ is again equal to $\frac{n-k}{n}$. However, if k is odd, $p_k \to (-1)^k = -1$. In this case

$$E(r_k) = -\frac{(n-k)}{n} - \frac{2Var(\overline{Z})}{\gamma_0}$$

$$= -\frac{(n-k)}{n} - \frac{2\Sigma\Sigma 1/n^2 \; Cov(Z_i, Z_j)}{\gamma_0}$$

$$= -\frac{(n-k)}{n} - \frac{2}{n^2}\Sigma\Sigma \; p_{|i-j|}$$

$$= -\frac{(n-k)}{n} \text{ if n is even}$$

$$= -\frac{(n-k)}{n} - 2/n^2 \text{ if n is odd .}$$

Hence, $E[S_1(k)] \approx 1/n-k$ and since $E[S_1(k)]$ has been shifted by a constant, the variance remains unchanged. Hence, K_1 is still valid as a test of $H_0 = \sigma_x^2 \leq 1$, and all that is required is a check of the average value of S_1 in order to determine whether $S_1(k) = 0$ or $S_1(k) = -2$, $k = 0, \pm 1, \pm 2, \ldots$.

B. TWO COMPLEX ROOTS ON UNIT CIRCLE

Suppose that the process

$$\phi(B)Z_t = \theta(B)a_t$$

has two complex roots approaching the unit circle. Then, Theorem 8 of [4] states that the autocorrelations p_k satisfy a second order difference equation with both roots complex and on the unit circle. Thus,

$$(3.4) \qquad (1-aB+B^2)p_k=0, \quad k = 0, \pm 1, \pm 2, \ldots,$$

which can be written

$$(1-G_1B)(1-G_2B)p_k=0, \; k = 0, \pm 1, \pm 2, \ldots ,$$

where

(3.5) $$G_1 G_2 = 1$$

(3.6) $$a = G_1 + G_2 = 2RE(G_1) = 2RE(G_2).$$

In particular, if $k = 1$, (3.4) implies

$$p_1 = ap_0 - p_1,$$

or

$$a = 2p_1,$$

so that (3.6) gives that

(3.7) $$p_1 = 1/2G_1 + 1/2G_2 .$$

Now, the general solution of (3.4) is given by

$$p_k = A_1 G_1^k + A_2 G_2^k,$$

but, with $p_0 = 1$, (3.7) implies that

$$A_1 + A_2 = 1$$

$$A_1 G_1 + A_2 G_2 = 1/2G_1 + 1/2G_2,$$

which immediately gives

$$A_1 = A_2 = 1/2,$$

$$p_k = 1/2G_1^k + 1/2G_2^k .$$

Note that

(3.8) $$2p_1 = G_1 + G_2,$$

which implies

$$G_2 = 2p_1 - G_1 ,$$

and if this is substituted into (3.5) results in

(3.9) $$G_1^2 - 2p_1G_1 + 1 = 0.$$

Solving (3.8) and (3.9) simultaneously for G_1 and G_2, then substituting into the general solution gives

(3.10) $$p_k = 1/2(p_1+\sqrt{p_1^2-1})^k + 1/2(p_1-\sqrt{p_1^2-1})^k.$$

By definitions (2.1) and (2.2),

$$S_2(k) = \frac{\begin{vmatrix} 1 & 1 & 1 \\ (-1)^k p_k & (-1)^{k+1} p_{k+1} & (-1)^{k+2} p_{k+2} \\ (-1)^{k+1} p_{k+1} & (-1)^{k+2} p_{k+2} & (-1)^{k+3} p_{k+3} \end{vmatrix}}{\begin{vmatrix} (-1)^k p_k & (-1)^{k+1} p_{k+1} \\ (-1)^{k+1} p_{k+1} & (-1)^{k+2} p_{k+2} \end{vmatrix}} .$$

Substitution of (3.10) into this definition and simplification produces the result

$$S_2(k) = 2p_1+2 \qquad k = 0,\pm 1,\pm 2,\dots ,$$

that is, the S_2 column is constant. Note also that

$$S_2(k) - 2 = 2p_1 = a.$$

Now, for each entry in the S_2 column, define

$$\tau_i = 1/2S_2(i) - 1.$$

Then, each τ_i is an estimate of p_1, which is a correlation coefficient. Therefore, assuming normal noise, Theorem B in [2], p. 225 states that

$$X_i = 1/2 \ \ln[\frac{1 + \tau_i}{1 - \tau_i}]$$

is normally distributed with

$$E(X_i) \approx 1/2 \ \ln[\frac{1+p_1}{1-p_1}] + \frac{p_1}{2n-1},$$

$$Var(X_i) \approx \frac{1}{n-3}.$$

Since τ_i estimates p_1 only under the assumption of two complex roots on the unit circle, it becomes necessary to have another estimate of p_1, independent of this assumption, in order to estimate $E(X_i)$. But, r_1 is the sample first lag autocorrelation, and, by definition,

$$-S_1(0)-1 = r_1.$$

Therefore, define

$$Z_i = \sqrt{n-3}(X_i - 1/2\ln[\frac{S_1(0)}{S_1(0)-2}] - \frac{S_1(0)+1}{2n-1}).$$

Then, the Z_i's constitute a sample from an approximate standard normal distribution. Again, to avoid effects from moving average operators of order three or less, define

$$I_{m'} = \{-m,\ldots,-5,2,\ldots,m-3\}.$$

Next, let

$$K_2 = \frac{k-1}{k} \sum_{i \varepsilon I_{m'}} (Z_i - \overline{Z})^2$$

where k is the number of elements in $I_{m'}$. Then, K_2 is distributed approximately as a chi-square with k-1 degrees of freedon, and testing if S_2 column is constant is equivalent to testing H_0: $\sigma_Z^2 \leq 1$, with the rule being to reject H_0 if

$$K_2 > X^2_{k-1,1-\alpha}.$$

An example from [4], the series generated for example 6, gives

i	$S_2(i)$	Z_i
-8	3.736	.671
-7	3.734	.611
-6	3.733	.581
-5	3.732	.552
2	3.690	-.626
3	3.690	-.626
4	3.689	-.641
5	3.688	-.671

Since $K_2 = 2.7$ and $X^2_{7,.95} = 14.1$, we assume an operator $(1\text{-}aB+B^2)$. Letting $C = 1/k \sum_{i \varepsilon I_{m'}} S_2(i) = 3.712$,

$$a = C-2 = 1.712,$$

and the operator we assume is

$$(1 - 1.712B + B^2).$$

C. TWO REAL ROOTS: +1 AND -1

Suppose the process

$$\phi(B)Z_t = \theta(B)a_t$$

has two real roots approaching the unit circle, one approaching positive unity and one approaching negative unity. Then, by Theorem 8 of [4], the autocorrelations p_k satisfy a second order difference equation with roots of plus and minus one. That is,

$$p_k = p_{k-2} \qquad k = 0, \pm 1, \pm 2, \ldots .$$

This implies that

$$p_k = 1 \text{ for } k \text{ even},$$

$$p_k = p_1 \text{ for } k \text{ odd},$$

and thus, (2.1) becomes

$$f_m = 1 \text{ for } m \text{ even},$$

$$f_m = -p_1 \text{ for } m \text{ odd}.$$

Next, it is noted that for m even,

$$S_2(m) = \frac{\begin{vmatrix} 1 & 1 & 1 \\ 1 & -p_1 & 1 \\ -p_1 & 1 & -p_1 \end{vmatrix}}{\begin{vmatrix} 1 & -p_1 \\ -p_1 & 1 \end{vmatrix}}$$

and for n odd,

$$S_2(m) = \frac{\begin{vmatrix} 1 & 1 & 1 \\ -p_1 & 1 & -p_1 \\ 1 & -p_1 & 1 \end{vmatrix}}{\begin{vmatrix} -p_1 & 1 \\ 1 & -p_1 \end{vmatrix}}$$

In each case, it is clear by inspection of the numerators that

$$S_2(m) = 0 \qquad m = 0, \pm 1, \pm 2, \ldots .$$

However, if the numerator of

$$S_2(m) = \frac{\begin{vmatrix} 1 & 1 & 1 \\ f_m & f_{m+1} & f_{m+2} \\ f_{m+1} & f_{m+2} & f_{m+3} \end{vmatrix}}{\begin{vmatrix} f_m & f_{m+1} \\ f_{m+1} & f_{m+2} \end{vmatrix}}$$

is expanded by the cofactors of the third column, it can be written as

(3.11)

$$S_2(m) = \frac{\begin{vmatrix} f_m & f_{m+1} \\ f_{m+1} & f_{m+2} \end{vmatrix} - f_{m+2} \begin{vmatrix} 1 & 1 \\ f_{m+1} & f_{m+2} \end{vmatrix} + f_{m+3} \begin{vmatrix} 1 & 1 \\ f_m & f_{m+1} \end{vmatrix}}{\begin{vmatrix} f_m & f_{m+1} \\ f_{m+1} & f_{m+2} \end{vmatrix}}.$$

Now, by definition,

$$S_1(m) = \frac{\begin{vmatrix} 1 & 1 \\ f_m & f_{m+1} \end{vmatrix}}{f_m}$$

which is identical to

$$\begin{vmatrix} 1 & 1 \\ f_m & f_{m+1} \end{vmatrix} = f_m S_1(m).$$

Substitution into (3.11) results in

$$S_2(m) = 1 + \frac{-f_{m+1}f_{m+2}S_1(m+1) + f_m f_{m+3} S_1(m)}{f_m f_{m+2} - f_{m+1}^2}$$

Under the supposed conditions of roots of plus and minus one, this reduces to

$$S_2(m) = 1 - \left(\frac{p_1}{1-p_1^2}\right)[S_1(m+1) - S_1(m)], \text{ n even,}$$

$$S_2(m) = 1 + \left(\frac{p_1}{1-p_1^2}\right)[S_1(m+1) - S_1(m)], \text{ m odd.}$$

Then, for any n,

$$\mathrm{Var}\{S_2(m)\} = \left(\frac{p_1}{1-p_1^2}\right)^2 \{\mathrm{Var}[S_1(m+1)] + \mathrm{Var}[S_1(m)]\}.$$

Now, the variance of the S_1 column under the assumption of one root equal to unity in absolute value has already been determined. Assumming that the variance of the S_1 remains the same as in the first order case as roots approach plus and minus one simultaneously, and assuming n large,

$$Var[S_2(m)] \approx (\frac{p_1}{1-p_1})^2 [\frac{4}{n-m} - \frac{2}{(n-m)^2}].$$

Since

$$E[S_2(m)] \approx 0,$$

and estimating p_1 as before with

$$r_1 = -S_1(0) - 1,$$

define

$$X_i = \frac{S_2(i)}{(\frac{r_1}{1-r_1^2})\sqrt{\frac{4}{n-i} - \frac{2}{(n-i)^2}}}.$$

Then, the X's correspond to a sample from an approximate standard normal distribution, so that

$$H_0: \sigma_x^2 \leq 1$$

can be tested by defining

$$K_3 = \frac{k-1}{k} \sum_{i \varepsilon I_{m'}} (X_i - \bar{X})^2$$

where k and $I_{m'}$ are the same as before. Then, K_3 is distributed approximately as a chi-square with k-1 degrees of freedom, and the rule may be stated as: Assume an operator $(1-B^2)$ unless $K_3 > X^2_{k-1,(1-\alpha)}$.

As an example, let the series be the series of example 9 in [4]. Then,

i	$S_2(i)$	X_i
-7	-.045	-1.356
-6	-.004	.120
-5	-.037	-1.108
2	.035	1.024
3	-.004	-.117
4	.042	1.221

Since $K_3 = 4.7$ and $X^2_{5,.95} = 11.1$, an operator $(1-B^2)$ is assumed.

IV. CONCLUSIONS

It was noted in [4] that the authors felt that some statistic similar to their D-statistic could be found that could be used to remove non-stationarities "rather automatically." The statistics K_1, K_2, and K_3 proposed in the thesis appear to provide such a method, and the program presented in the thesis makes use of K_1, K_2, and K_3 for that purpose. Unlike the D-statistic, K_1, K_2, and K_3 are such that confidence levels can accompany the decisions. However, the proposed statistics work only for particular cases, and not in general. Also, moving average operators present problems in the effectiveness of K_1, K_2, and K_3. Nevertheless, in the restricted class of third order models and second order non-stationarities, the thesis presents and implements a practical solution to the problem of identifying and analyzing non-stationary, as well as stationary, time series.

The FORTRAN-IV program to analyze a time series is available in card format from Dr. Thomas L. Boullion, Department of Mathematics, Texas Tech University, Lubbock, Texas. Included are explanations of print control parameters and input formats.

REFERENCES

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

2. Brunk, H.D., (1975), An Introduction to Mathematical Statistics, Lexington, Mass: Xerox.

3. Fuller, Wayne A., (1976), Introduction to Statistical Time Series, New York: John Wiley & Sons.

4. Gray, Kelley, & McIntire, (1978), A New Approach to ARMA Modeling: Comm. in Stat. B pp. 1-77.

5. Mood, Graybill, and Boes, (1974), Introduction to the Theory of Statistics: New York: McGreer-Hill.

6. Nelson, David Lauren,(1969),"Numerical Methods for the Solution of Non-Linear Least Squares Problems", Master's Thesis, Texas Tech.