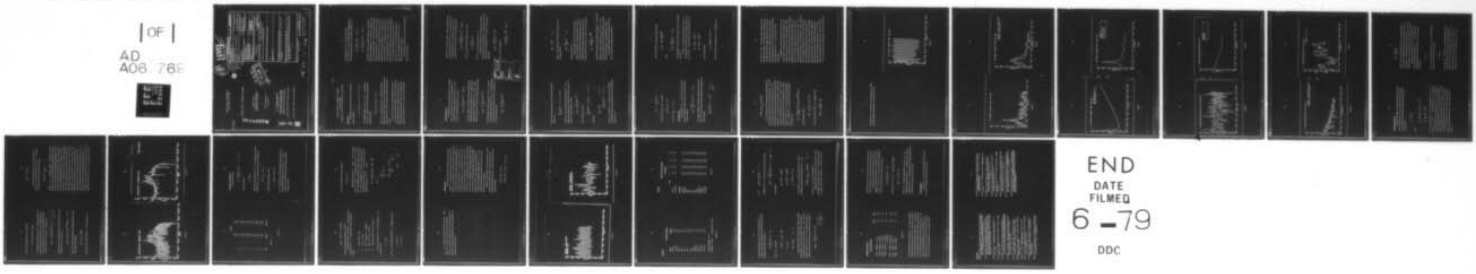
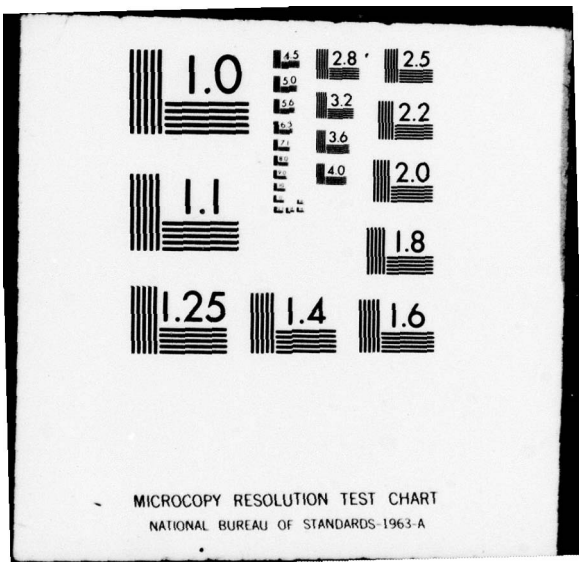
AD-A067 769 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 12/1
SOME RECENT ADVANCES IN AUTOREGRESSIVE PROCESSES.(U)
FEB 79 M PAGANO N00014-78-C-0599
UNCLASSIFIED TR-N-3 NL



END
DATE FILMED
6-79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843

INSTITUTE OF STATISTICS
Phone 713 - 845-3141

SOME RECENT ADVANCES IN AUTOREGRESSIVE PROCESSES

By Marcello Pagano

Department of Biostatistics
Harvard School of Public Health

Technical Report No. N-3

February 1979

Texas A & M Research Foundation
Project No. 3838

"Multiple Time Series Modeling and Time Series Theoretic Statistical Methods"

Sponsored by the Office of Naval Research

Professor Emanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ DISTRIBUTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report No. N-3
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Some Recent Advances in Autoregressive Processes
5. TYPE OF REPORT & PERIOD COVERED: Technical
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Marcello Pagano
8. CONTRACT OR GRANT NUMBER(s): N00014-78-C-0599
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Texas A&M University, Institute of Statistics, College Station, TX 77843
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Code 436, Arlington, VA 22217
12. REPORT DATE: February 1979
13. NUMBER OF PAGES: 42
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 22p.
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report): TR-N-3
Approved for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report): NA
18. SUPPLEMENTARY NOTES: NA
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): Autoregressive schemes, seasonal autoregression, autoregressive order determination
20. ABSTRACT (Continue on reverse side if necessary and identify by block number): A survey paper presented at the Special Topics Meeting on Time Series Analysis, organized by the Institute of Mathematical Statistics at Ames, Iowa, May 1-3, 1978.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SOME RECENT ADVANCES IN AUTOREGRESSIVE PROCESSES

By Marcello Pagano

Department of Biostatistics, Harvard School of Public Health

Introduction

Motivated by the fact that the solution of the second order difference equation,

$$z(t) - 2\cos w\, z(t-1) + z(t-2) = 0\,, \quad t = 2, 3, \ldots$$

with starting conditions, $z(1) = \cos w$, $z(0) = 1$, is the sinusoid, $z(t) = \cos wt$, Udny Yule in 1927 proposed the (second order) autoregressive process,

$$Y(t) + a_1 Y(t-1) + a_2 Y(t-2) = \epsilon(t)\,,$$

where $\epsilon(\cdot)$ is a white noise process, as an alternative to the then dominant purely spectral domain approach to modeling periodic phenomena. Thus the birth of the, so called, "time domain" approach to time series analysis.

This was also the start of an unfortunate dichotomy of time series methods into, with few exceptions, exclusive, time domain and spectral domain approaches. Unfortunate, since results in either domain usually may be interpreted in the other, thus lending another dimension in the analysis of series. Given a zero mean stationary

(discrete) time series $Y(\cdot)$ with covariance function $R(v) = EY(t)\,Y(t+v)$, then the pivotal quantity for linear methods is either the spectral density $f(\cdot)$ or its Fourier coefficients $R(\cdot)$;

$$R(v) = \int_{-\pi}^{\pi} f(w)\, e^{ivw}\, dw\,, \quad v = 0, \pm 1, \ldots$$

$$f(w) \sim \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} R(v)\, e^{-ivw}\,, \quad w \in [-\pi, \pi]\,.$$

The problem at hand is to estimate either function. Henceforth we make the mild assumption that $f(\cdot)$ is positive, or equivalently, that $R(\cdot)$ is strictly positive definite. Time domain methods are usually concerned with modeling $R(\cdot)$, and unfortunately the raw estimator is problematic since it is correlated. Whereas, frequency domain methods are concerned with modeling $f(\cdot)$, and the problem here is that the raw estimator is not consistent thus requiring that some form of smoothing be performed. In either case, because of only a finite amount of data, there are two errors to be made, and optimally, to be minimized; the approximation error, at best only a finite number of the $R(\cdot)$, or $f(\cdot)$ at a finite number of points, are known, and, the estimation error, neither $R(\cdot)$ nor $f(\cdot)$ is usually known and must be estimated.

We concentrate on showing how the autoregressive approximant fits into the above schema.

Approximation Error

Concentrating first on the approximation error, assume that $R(0), \ldots, R(p)$ are given, and that we wish to approximate $f(\cdot)$. The Riemann-Lebesgue lemma states that $R(v) \to 0$, as $|v| \to \infty$, and thus a reasonable approximant would seem to be

$$f_p(\omega) = \frac{1}{2\pi} \sum_{v=-p}^{p} R(v)\, e^{-iv\omega} , \quad \omega \in (-\pi, \pi] .$$

Thus $f_p(\cdot)$ matches $f(\cdot)$ in its first p Fourier coefficients and replaces the higher order coefficients by zero. Because of the undesirable convergence properties (as $p \to \infty$) of $f_p(\cdot)$, Fejér introduced, what came to be called, the windowed spectral estimator,

$$f_{p,w}(\omega) = \frac{1}{2\pi} \sum_{v=-p}^{p} w(v)\, R(v)\, e^{-iv\omega} ,$$

which has superior convergence properties. If the $w(\cdot)$ sequence is positive definite, thus ensuring that $f_{p,w}(\cdot)$ is nonnegative (surely a desirable characteristic, because of its rolè as an approximant for a positive function), then

$$f_{p,w}(\omega) = \frac{1}{2\pi} \sum_{v=-p}^{p} R^*(v)\, e^{-iv\omega}$$

where $R^*(\cdot)$ is a covariance function. Indeed, even though arrived at purely in the spectral domain, $f_{p,w}(\cdot)$ may be called a moving average approximant since it is the spectral density of a moving average process: we have from the Fejér-Riesz theorem that

$$f_{p,w}(\omega) = \frac{\sigma^2}{2\pi} \left| \sum_{k=0}^{p} \beta(k)\, e^{-ik\omega} \right|^2 , \quad \beta(0) = 1$$

so that $f_{p,w}(\cdot)$ is the spectral density of $X(\cdot)$ where,

$$X(t) = \sum_{k=0}^{p} \beta(k)\, \epsilon(t-k)$$

and $\epsilon(\cdot)$ is white noise with variance σ^2.

An alternative approximant is the autoregressive approximant. Whereas the moving average approximant is a trigonometric polynomial, the autoregressive approximant is the reciprocal of a trigonometric polynomial;

$$f_{p,a}(\omega) = \left(2\pi \sum_{v=-p}^{p} Ri(v)\, e^{-iv\omega} \right)^{-1}$$

which, since it is nonnegative,

$$= \sigma_p^2 / 2\pi \left| \sum_{j=0}^{p} \alpha_p(j)\, e^{-ij\omega} \right|^2 , \quad \alpha_p(0) = 1 .$$

This is the spectral density function of an autoregressive process (of order p) .

$$\sum_{j=0}^{p} a_p(j)\, X(t-j) = \epsilon(t) \quad , \quad t = 0, \pm 1, \ldots ,$$

where the $\epsilon(\cdot)$ process is white noise with variance σ^2 . The autoregressive approximant is usually chosen so as to match $f(\cdot)$ in its first p Fourier coefficients: letting

$$f_{p,a}(\omega) = \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} R_p(v)\, e^{-iv\omega}$$

then if we define the $a_p(\cdot)$ by

$$\sum_{j=0}^{p} a_p(j)\, R(v-j) = \sigma_p^2\, \delta_{v,0} \qquad v = 0, \ldots, p$$

we have that

$$R_p(v) = R(v) \qquad v = 0, \ldots, p$$

$$= -\sum_{j=1}^{p} a_p(j)\, R_p(v-j) \qquad v = p+1, \ldots \quad .$$

So that $f_{p,a}(\cdot)$ matches $f(\cdot)$ in its first p Fourier coefficients and replaces the higher order coefficients with numbers exponentially decaying to zero:

$$R_p(v) = 0(\rho^{-|v|}) \quad , \qquad \rho > 1 \quad ,$$

indeed, ρ is the modulus of the zero (all outside the unit circle, see Pagano (1973)) of

$$g_p(z) = \sum_{j=0}^{p} a_p(j)\, z^j \quad , \qquad z \in \mathbb{C} \quad ,$$

closest to the unit circle.

It may seem unnatural to use the reciprocal of the polynomial as an approximant, rather than a polynomial, but there is a large body of evidence that it yields good practical results, and, secondly, Masani (1966) gives conditions under which a general spectral density may be written as an infinite order autoregression,

$$f(\omega) = \sigma^2/2\pi \left| \sum_{j=0}^{\infty} a_\infty(j)\, e^{-ij\omega} \right|^2$$

(certainly true under the mild condition that $\sum_{v=-\infty}^{\infty} |R(v)| < \infty$). Kromer (1969) studies the convergence properties of $f_{p,a}(\cdot)$ (as $p \to \infty$) to $f(\cdot)$. One should also mention the work of Carmichael (1978), who uses the autoregressive approximant as a probability density estimator.

One immediate advantage of the autoregressive approximant is that it yields the best memory-p linear predictor: if

$$Y(t / t-1) = \sum_{j=1}^{p} \lambda(j)\, Y(t-j)$$

and we wish to find the $\lambda(\cdot)$ to minimize

$$E[Y(t) - Y(t / t-1)]^2 ,$$

then $\lambda(j) = -a_p(j)$, $j = 1, \ldots, p$.

Estimation Error

Given a sample $Y(1), \ldots, Y(T)$ from a zero mean and, for the sake of definiteness, Gaussian time series, define the sample covariance function

$$R_T(v) = \sum_{t=1}^{T-|v|} Y(t)\, Y(t + |v|) , \quad v = 0, \pm 1, \ldots .$$

The Gaussian assumption can easily be relaxed (see Hannan (1970)). Then a p^{th} order autoregressive approximant can be estimated via the normal equations

$$\sum_{j=0}^{p} a_p(j)\, R_T(v-j) = \delta_{v,0}\, \hat{\sigma}_p^2 \qquad v = 0, \ldots, p$$

$$a_p(0) = 1 .$$

Whence the estimator of the spectral density estimator,

$$\hat{f}_{p,a}(\omega) = \hat{\sigma}_p^2 / 2\pi \left| \hat{g}_p(e^{i\omega}) \right|^2$$

where

$$\hat{g}_p(z) = \sum_{j=0}^{p} a_p(j)\, z^j , \qquad z \in \mathbb{C}$$

and is guaranteed to have its zeroes outside the unit circle (unless all the $Y(t) = 0$).

When the $Y(\cdot)$ process is an autoregression of order p , then Mann and Wald (1943) determined the sampling properties of the $a_p(\cdot)$ and $\hat{\sigma}_p^2$. Indeed, they show that if we define the vectors

$$\underset{\sim}{a}_p = (a_p(1), \ldots, a_p(p))^T ,$$

$$\underset{\sim}{\alpha}_p = (\alpha_p(1), \ldots, \alpha_p(p))^T ,$$

and the Toeplitz matrix

$$\underset{\sim}{T}_p = \text{Toepl}(R(0), \ldots, R(p-1)) ,$$

then $\underset{\sim}{a}_p$ and $\hat{\sigma}_p^2$ are asymptotically $(T \to \infty)$ independent and

$$L(\sqrt{T}(\hat{a}_p - a_p)) \underset{T\to\infty}{\to} N(0, \sigma_p^2 \Gamma_p^{-1}) \quad .$$

The more difficult problem of deciding the sampling distribution of the autoregressive approximant when the sample is not from a p^{th} order autoregression, is handled by Kromer (1969) and Berk (1974). They obtain results which are asymptotically equivalent to the moving average approximant. One result of this work is a solution of a problem we have not addressed, namely deciding on the order p to choose. We see the conflicting considerations made by the approximation error, which is minimized by large p $(p = \infty)$, whereas variance arguments support a small p. Parzen (1976), as a compromise suggests the integrated mean square error

$$M(p) = \int_{-\pi}^{\pi} E\left|\frac{1}{\hat{\sigma}_p^2}\hat{g}_p(e^{i\omega}) - \frac{1}{\sigma_\infty^2} g_\infty(e^{i\omega})\right|^2 f(\omega)\, d\omega \quad ,$$

and shows that for large T,

$$M(p) = \frac{1}{T}\sum_{j=1}^{p} \sigma_j^{-2} + \sigma_\infty^{-2} - \sigma_p^{-2} \quad .$$

Thus as an estimator of $M(p) - \sigma_\infty^{-2}$ he suggests we use

$$CAT(p) = \frac{1}{T}\sum_{j=1}^{p} \tilde{\sigma}_j^{-2} - \tilde{\sigma}_p^{-2}$$

with $\tilde{\sigma}_j^2 = (T/T - j)\hat{\sigma}_j^2$. The order to be fit can be chosen as that order which minimizes CAT. Other approaches for deciding the order to fit include the criterion of Akaike (1972) (which yields results similar to CAT) and the more classical approaches covered in Anderson (1972) and Hannan (1970). These last methods are somewhat dependent on there being a finite order p; i.e., the sample is from an autoregression.

Our experience with CAT is that it yields good answers. One example, Figure 1, shows the average monthly temperature for New York City for the years 1946 - 1959. The log periodogram is shown in Figure 2 and in Figure 3 we have the log of the periodogram smoothed with the Parzen window with three different truncation points (16, 32 and 64). The CAT criterion (Figure 4) chose order 9 for the autoregressive approximant, the log of which is shown in Figure 5. The expected peak at 12 months is very well defined as are the main harmonics. Other such examples could be shown, but I would rather show the one series which indicates that further work needs to be done with CAT. Figure 6 shows the 720 hourly readings of the azimuth of a pendulum (Allais (1961)) and Figure 7 shows the autoregressive approximant. This seems too flat in light of the smoothed spectral estimators in Figure 8 (Parzen window with truncation points 72, 144 and 288). Further, the residual variance estimator $\tilde{\sigma}_j^2$ reaches a minimum at $j = 38$, as shown in Figure 9, which would tend to indicate that CAT chose

too small an order. This seems to be an example where one should use subset autoregressive methods, such as in McClave (1976).

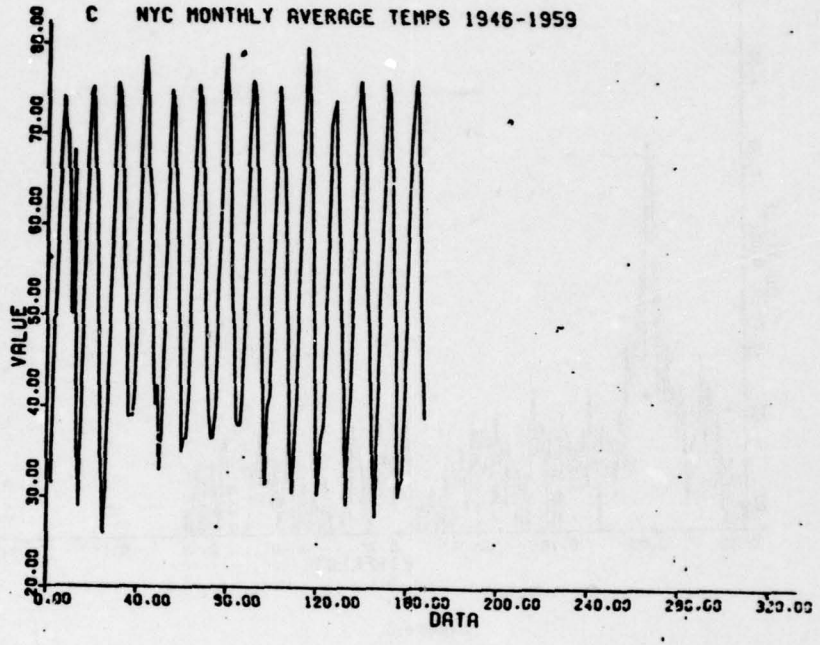


Figure 1.

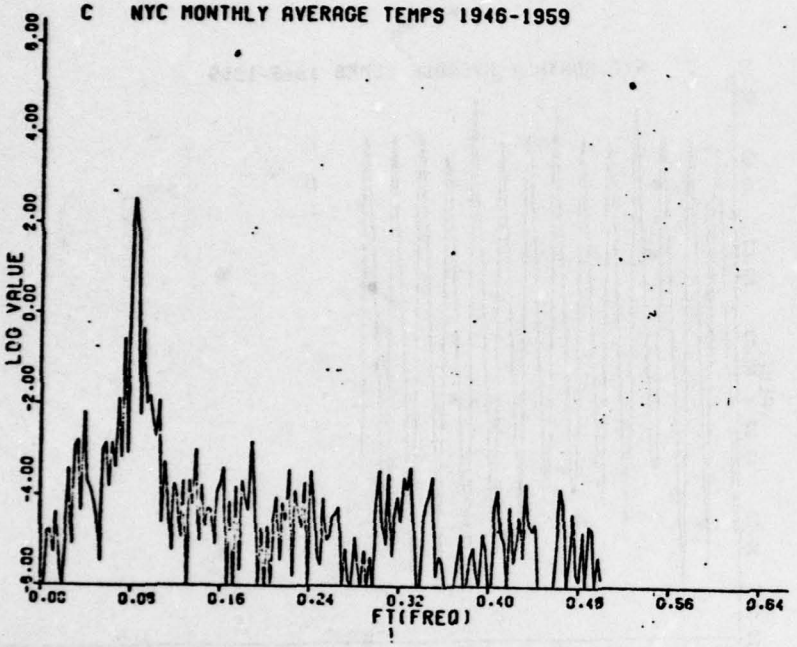


Figure 2.

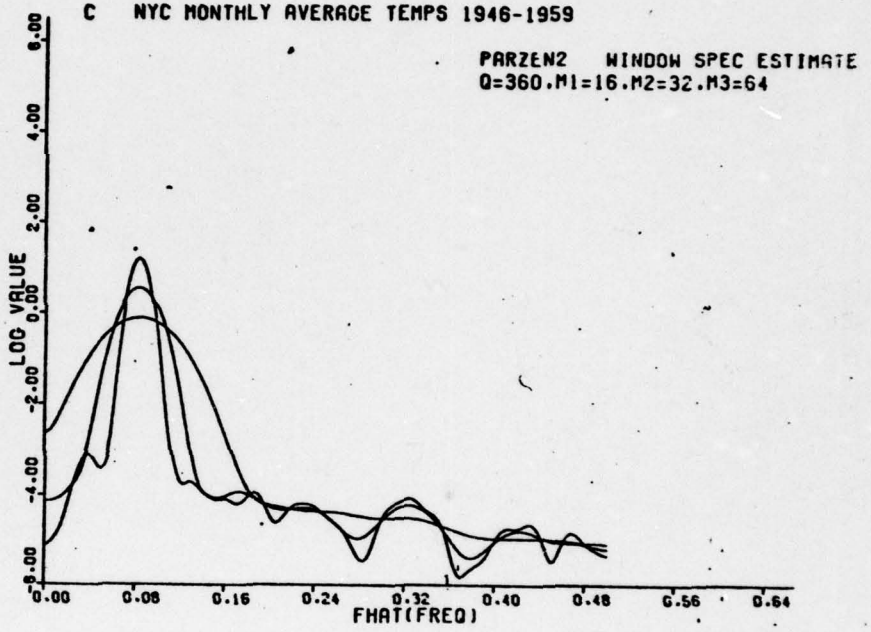


Figure 3.

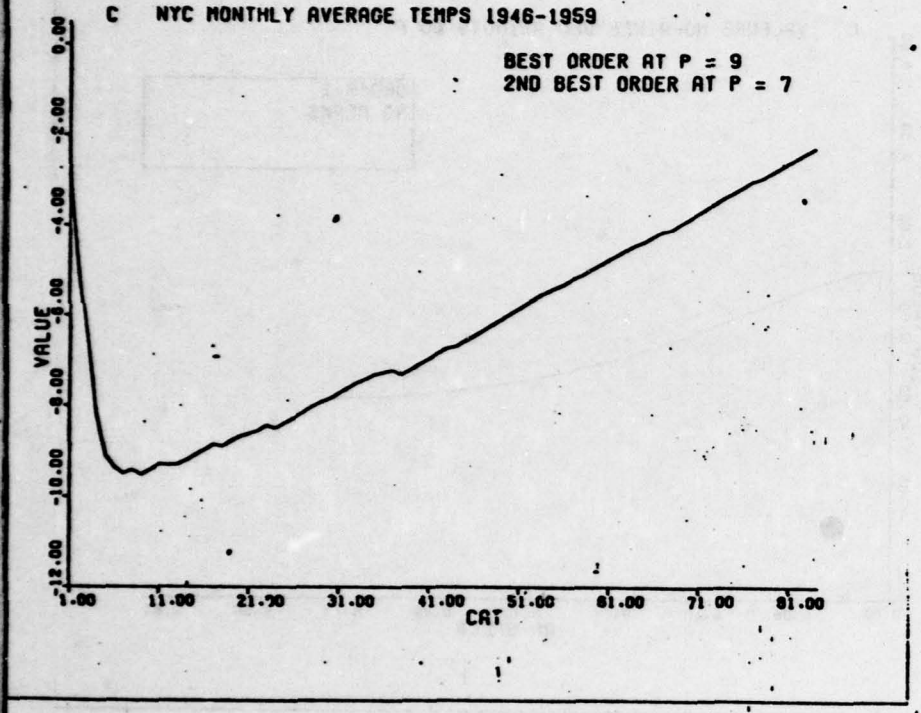


Figure 4.

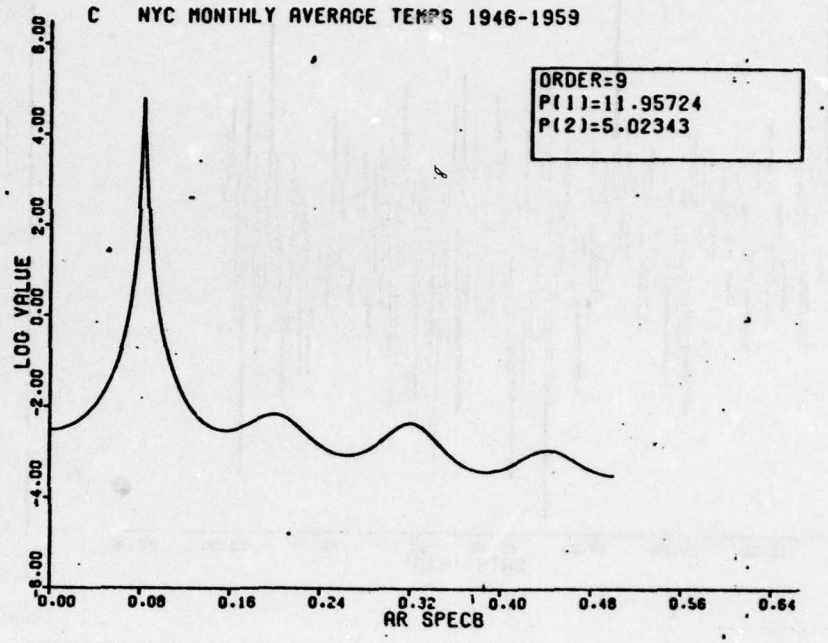


Figure 5.

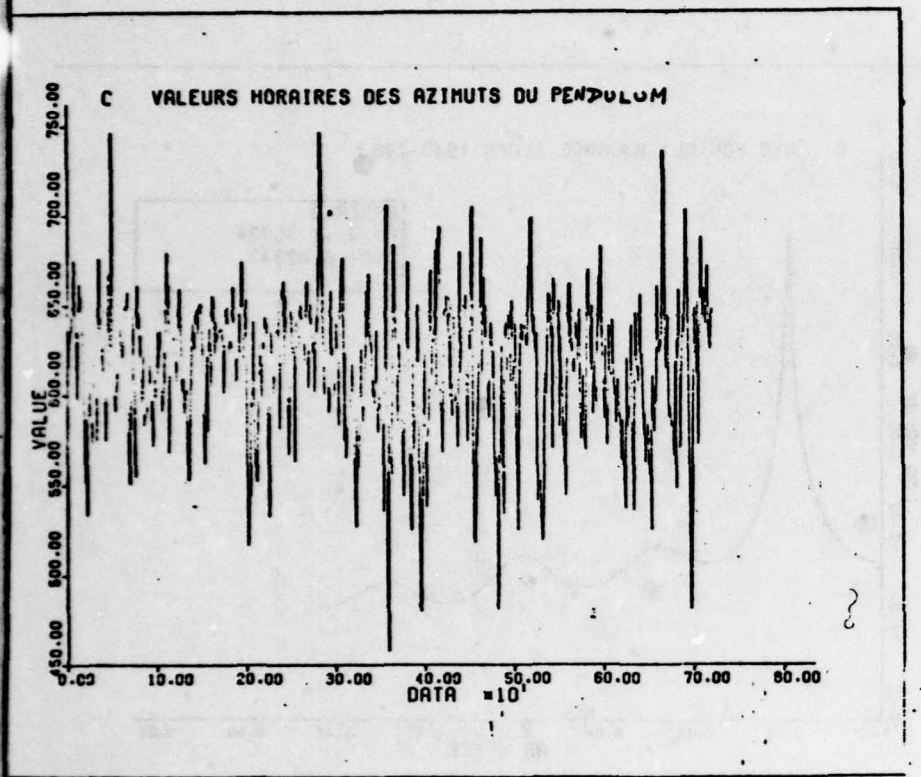


Figure 6.

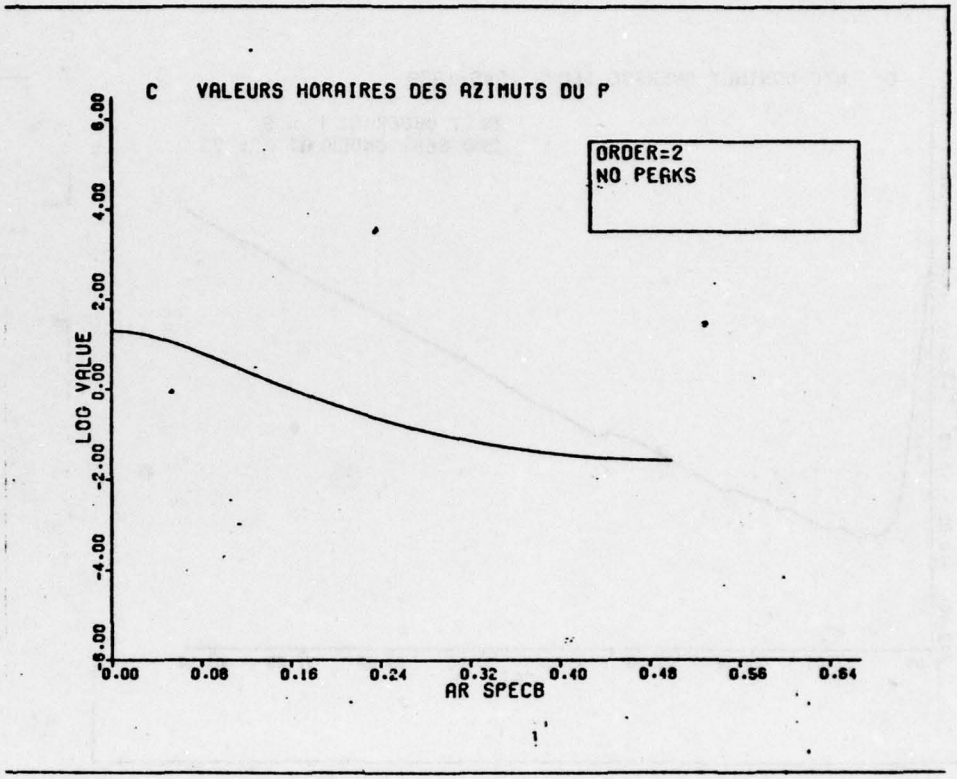


Figure 7.

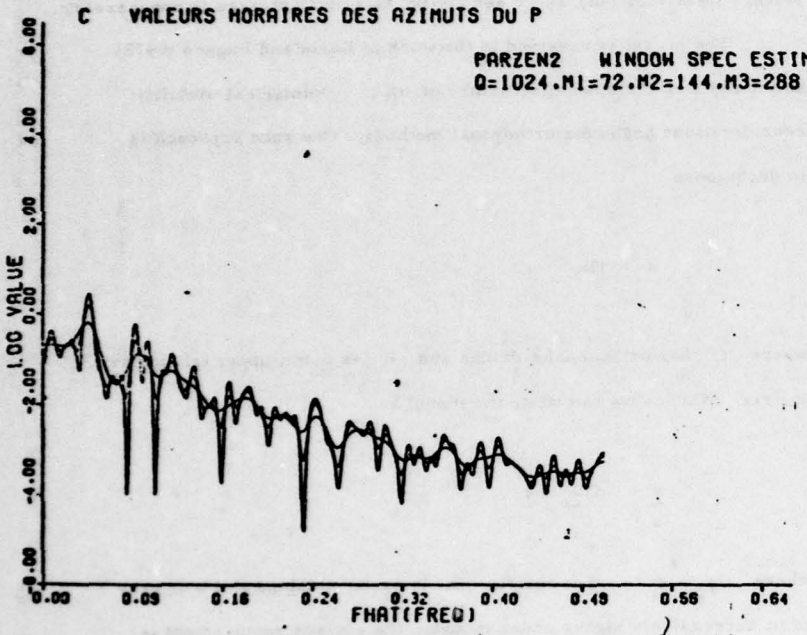


Figure 8.

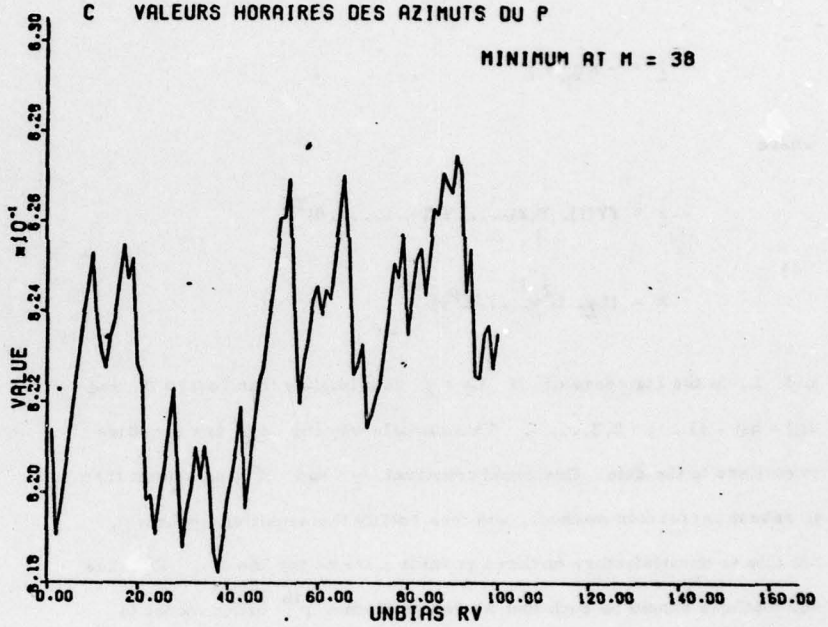


Figure 9.

Fitting Autoregressions

A study of the algorithms used for fitting autoregressions reveal some interesting results. The $\underset{\sim}{a}_p$ obtained in the previous sections are the estimators one would obtain by using least squares methods on the linear model,

$$\underset{\sim}{y} = -X\underset{\sim}{a}_p + \underset{\sim}{\varepsilon}$$

where

$$\underset{\sim}{y} = (Y(1), Y(2), \ldots, Y(T), 0, \ldots, 0)^T$$

$$X = (L\underset{\sim}{y}, L^2\underset{\sim}{y}, \ldots, L^p\underset{\sim}{y})$$

and L is the lag operator: if $L\underset{\sim}{a} = \underset{\sim}{b}$ this implies that $a(1) = 0$ and $a(j) = b(j-1)$, $j = 2, 3, \ldots$. This reveals why the $\underset{\sim}{a}_p$ are sensitive to outliers in the data. One could construct $\underset{\sim}{y}$ and X and submit them to robust regression methods, and thus rectify the sensitivity problem, but this is unsatisfactory on three grounds: we do not know p and thus our methods should be such that we can go from a p^{th} order model to a $(p+1)^{th}$ order model easily. Ordinary robust regression methods would not have this facility. Secondly, we require our filter to be stable, i.e.

$$\hat{g}(z) = \sum_{j=0}^{p} a_p(j)\, z^j$$

must have all its zeroes outside the unit circle; a difficult constraint to apply. Thirdly, only moderate storage requirement should be made of the computer. As it stands it seems that we would require $(p+1) \times (T+p)$ storage locations, but, as we see below, this much storage is unnecessary.

The secret is revealed in the work of Davis and Pagano (1978) and it rests on the special structure of X. Numerical stability considerations argue for orthogonal methods. One such approach is to decompose

$$X = QR$$

where Q has orthogonal columns and R is a unit upper triangular matrix. Whence we can write the model as

$$\underset{\sim}{y} = -Q\underset{\sim}{\theta}_p + \underset{\sim}{\varepsilon}$$

where $R\underset{\sim}{a}_p = \underset{\sim}{\theta}_p$. The matrix Q is easily obtained and it is easy to fit successively higher order models, the storage requirement is small, $2T$ (unless the diagonal elements of $QQ^T = X(X^TX)^{-1}X^T$ are also required, in which case it is obviously $3T$). The stability condition is translated to the easily applied condition that $|\theta_p(j)| < 1$.

This is not surprising since it can be shown that

$$\underset{\sim}{\alpha}_p^T = (\alpha_1(1), \alpha_2(2), \ldots, \alpha_p(p)) \ ,$$

i.e., the partial autocorrelations. These considerations would tend to argue that the better parametrization is $\underset{\sim}{\alpha}_p$ rather than the $\underset{\sim}{a}_p$. One can easily obtain the latter from the former using the Levinson recursions (making the calculation of R unnecessary)

$$a_m(j) = a_{m-1}(j) + a_m(m)\, a_{m-1}(m-j) \qquad \begin{matrix} j = 1, \ldots, m \\ m = 1, \ldots, p \end{matrix}$$

which are a special case of the Szego recursions which even give us the spectral density in terms of the $\underset{\sim}{\alpha}_p$:

$$g_m(\omega) = g_{m-1}(\omega) + a_m(m)\, e^{i\omega m} \overline{g}_{m-1}(\omega) \qquad m = 1, 2, \ldots \ .$$

Further, the partial autocorrelations have the usual reduction in innovation variance interpretation,

$$\sigma_m^2 = \sigma_{m-1}^2 (1 - \alpha_m^2(m)) \ ,$$

and they are the correlations between $\underset{\sim}{y}$ and the columns of Q ,

$$\alpha_m(m) = \text{corr}\,(\underset{\sim}{y}, \underset{\sim}{q}_m) \ .$$

This last interpretation allows us to use other than least squares as a criterion for estimating the $\alpha_m(m)$. For example, one could use the least absolute deviations criterion,

$$\min_a \sum_t |y(t) - a\, q_m(t)|$$

subject to $|a| < 1$, to estimate $\alpha_m(m)$. This criterion is particularly easy to implement using Boscovich's algorithm (1760). This was done for the data in Figure 10 which represents daily data for one year of electrical consumption for a large power company. The seven-day cycle is removed and large outliers are evident (Figure 11), due to public holidays. The two columns in Figure 12 represent the calculated partial correlations; the first column the least squares results and the second column the least absolute deviation results. The CAT criterion chooses order 6 on the first and order 1 on the second column. The big difference is due to the outliers. One could argue that the order 1 yields a better answer since it says that today's deviation from the mean is best predicted by yesterday's deviation from its mean $(.921 \approx 1)$. This seems sensible since the overriding determinant of electrical consumption, as measured in deviations from a daily mean, is the weather, and what better predictor of today's weather than yesterday's weather?

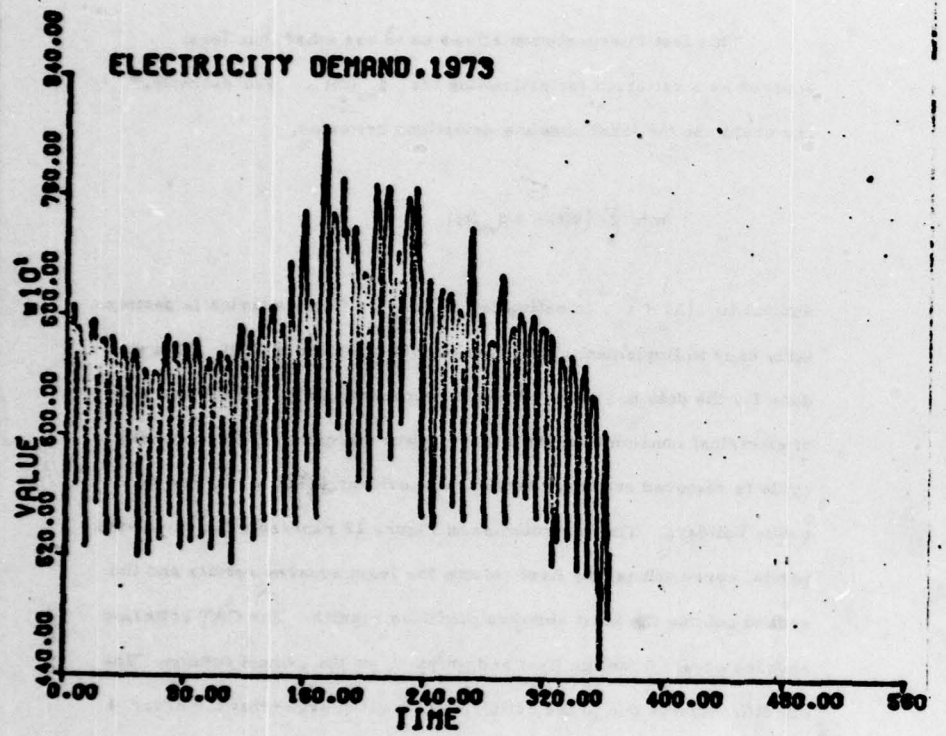


Figure 10.

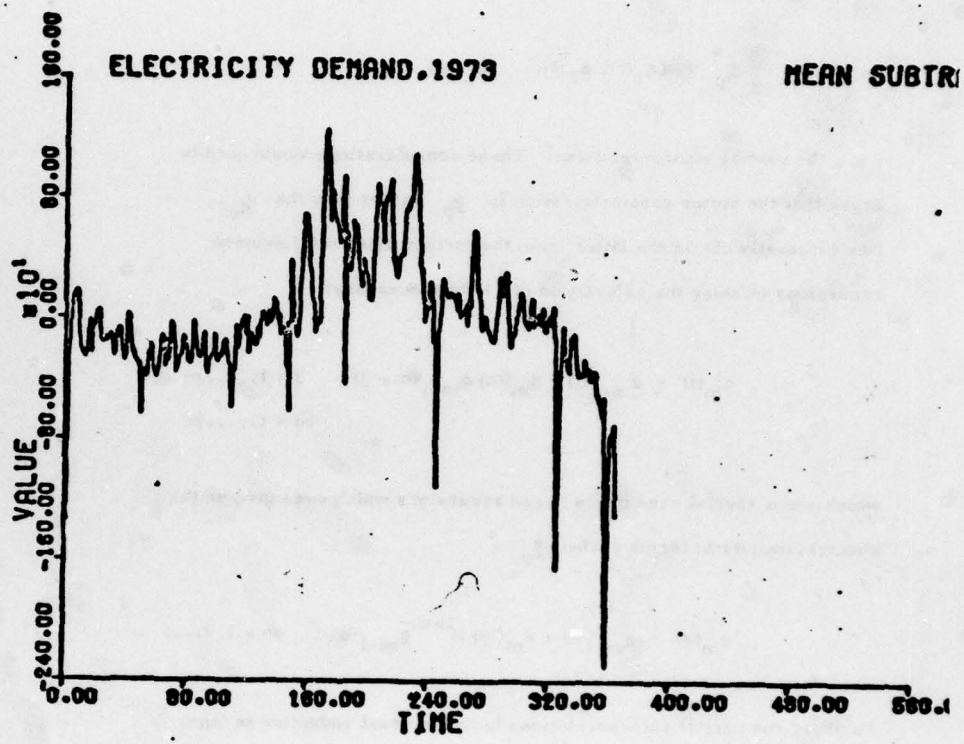


Figure 11.

$a_j(j)$

	ℓ_2	ℓ_1
1	-.820	-.921
2	-.175	-.047
3	-.056	+.002
4	-.052	-.006
5	-.084	-.077
6	-.128	-.055
7	.026	-.044
8	-.002	0.000
9	-.046	-.034
10	-.055	-.038
CAT:	ORDER 6	ORDER 1

$2/\sqrt{T} \approx .1$

Figure 12.

Residuals from Regression

Given the linear model

$$y(t) = \sum_{j=1}^{m} \beta(j)\, x_{t,j} + \eta(t) \quad , \qquad t = 1, \ldots, T$$

or in matrix notation,

$$\underline{y} = X\underline{\beta} + \underline{\eta}$$

where the covariance matrix of $\underline{\eta}$ is V, the weighted least squares estimator of $\underline{\beta}$ is given by the solution of the normal equations,

$$X^T V^{-1} X \underline{b} = X^T V^{-1} \underline{y} \quad ,$$

and its covariance matrix is given by

$$\text{Cov}(\underline{b}) = (X^T V^{-1} X)^{-1} \quad .$$

If V is unknown, one could use the ordinary least squares or any other estimator which ignores the covariance in $\underline{\eta}$. Often one does not err too much by proceeding in this manner but then the covariance matrix is

$$\text{Cov}(\underline{b}) = (X^T X)^{-1} X^T V^{-1} X (X^T X)^{-1}$$

which is usually poorly approximated by $(X^T X)^{-1}$.

An alternative approach is to estimate V and then proceed as if it were the correct V . For example one could use the following scheme:

1° Solve $X^T X \underline{b} = X^T \underline{y}$.

2° Let $\hat{\underline{\eta}} = \underline{y} - X\underline{b}$.

3° Use $\hat{\underline{\eta}}$ to estimate V yielding $\hat{V}$.

4° Solve $X^T \hat{V}^{-1} X \underline{b} = X^T \hat{V}^{-1} \underline{y}$.

5° Unless convergence is reached, return to 2° .

One way of performing step 4° is to first decompose V into its Cholesky factors

$$V = L D^2 L^T$$

where L is a unit lower triangular and D is diagonal. Then let

$$\underline{z} = D^{-1} L^{-1} \underline{y}$$

$$Z = D^{-1} L^{-1} X$$

and $\underline{b}$ is the solution to the normal equations

$$Z^T Z \underline{b} = Z^T \underline{z} .$$

The important step is step 3° . Clearly V has to be modeled since it has $1/2T(T+1)$ parameters and there are only T observations. One model that has proved to be useful is when we can model the $\eta(\cdot)$ series as a stationary series. Using an autoregressive approximant to estimate the covariance of $\eta(\cdot)$ has the further advantage of yielding very simple D and L^{-1} . Indeed,

$$\hat{D}^2 = \text{diag}\,(R_T(0), \theta_1^2, \ldots, \theta_p^2 \ldots, \theta_p^2)$$

$$\hat{L}^{-1} = \begin{pmatrix} 1 & & & & \\ a_1(1) & 1 & & & \\ a_2(2) & a_2(1) & 1 & & \phi \\ \vdots & & \ddots & & \\ a_p(p) & a_p(p-1) & \cdots & a_p(1) & 1 \\ & a_p(p) & \cdots & a_p(1) & 1 \\ & \ddots & & \ddots & \ddots \\ \phi & & a_p(p) \cdots & a_p(1) & 1 \end{pmatrix}$$

i.e., L^{-1} is a band lower triangular matrix with, at most, $1/2p(p-1)$ distinct nonzero elements. This fact follows from Levinson's algorithm. Oberhofer and Kmenta (1974) have studied the theoretical properties of such iteratively reweighted least squares algorithms, see also Pagano and Hartley (1975).

Seasonal Series

Often one observes and must model series such as displayed in Figure 13. This particular series represents the average monthly total ozone in the stratosphere above Tateno in Japan for the years 1958 to 1975. Figure 14 shows the residuals obtained when one subtracts the monthly means from Figure 13. It seems like there might still be some seasonal behaviour. If one calculates the monthly variances one obtains the numbers in Figure 15. One notices a ratio of 16 : 1 for the largest to the smallest monthly variances. This casts suspicion on the assumption of stationarity. If one investigates further (Figure 16) one sees that the correlation between Januaries and the immediately preceding Decembers is .688 whereas the correlation between Novembers and immediately preceding Octobers is -.405 . If the assumption of stationarity were valid these two numbers would be estimates of the same quantity.

Gladyshev (1961) introduced a class of models which are wider than the class of stationary models. He defines a process to be periodically stationary of period d if both its mean and covariance kernel are periodic;

$$E\,Y(t) = m(t) = m(t+d) \quad ,$$

$$\mathrm{Cov}\,(Y(t), Y(s)) = R(s,t) = R(s+d, t+d) \quad .$$

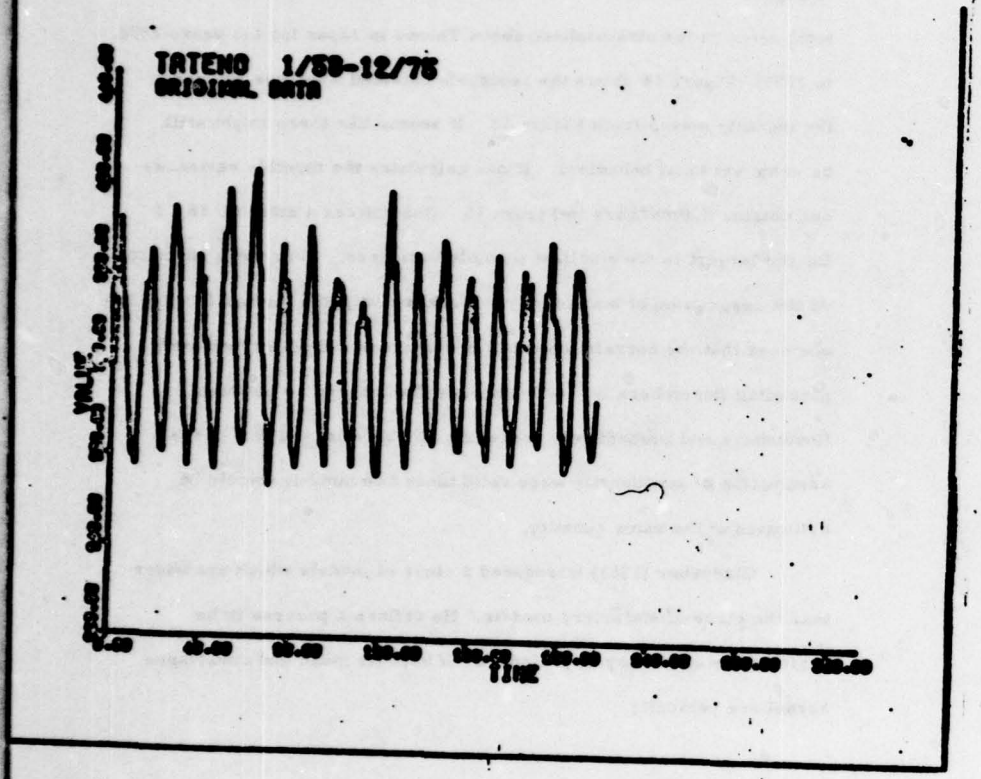


Figure 13.

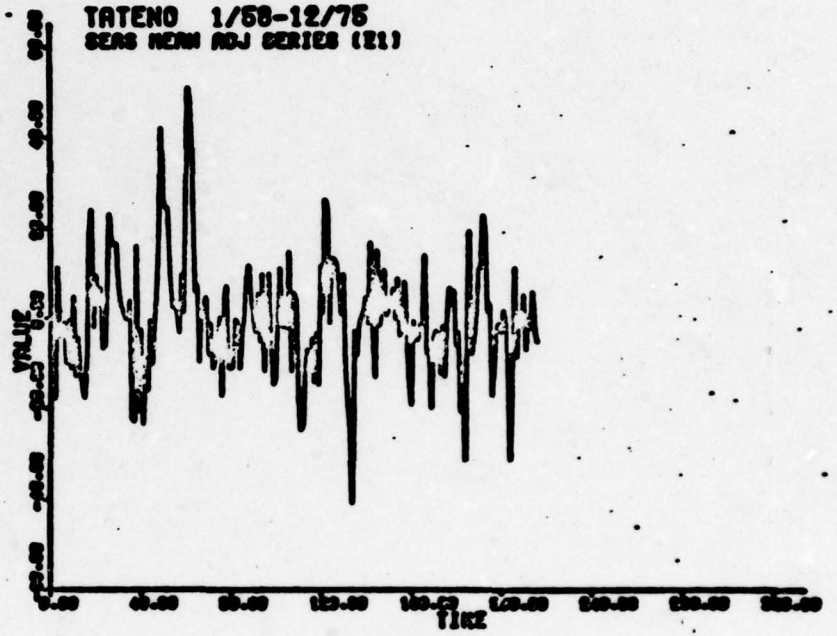


Figure 14.

TATENO

Seasonal Variances

Month	Variance
January	429.5
February	458.3
March	287.4
April	232.7
May	113.4
June	100.6
July	69.0
August	27.9
September	42.1
October	82.8
November	72.1
December	198.2

VARMAX/VARMIN = 16.4

(Bartlett's = 66.04 Critical Value = 17.28)

Figure 15.

TATENO

Seasonally Mean Adjusted

Correlogram

Lag:	1	2	3	4
Overall	.494	.365	.196	.100
Studentized	.345	.370	.156	.139
By Month				
January	.688	.174	.153	-.251
February	.677	.525	.191	-.125
March	.646	.567	.509	.145
April	.644	.567	.386	.231
May	.253	.525	.255	.424
June	.640	-.276	-.077	-.228
July	.119	.589	.560	.537
August	.464	.279	.409	.380
September	.337	.580	.127	.302
October	-.166	.400	-.040	.502
November	-.405	.504	-.362	.041
December	.239	.003	-.238	-.291

Figure 16.

A possible approximant for the second order properties of such a process is the periodic autoregression (see Jones and Brelsford (1967) and Pagano (1978)) which is defined by

$$Y(t) + \sum_{j=1}^{P_t} a_t(j)\, Y(t-j) = \epsilon(t)$$

where $\epsilon(\cdot)$ is noise with

$$E\,\epsilon(t)\,\epsilon(s) = \delta_{s,t}\,\sigma_t^2 = \sigma_{t+d}^2$$

$$P_t = P_{t+d}$$

$$a_t(j) = a_{t+d}(j) \qquad j = 1, \ldots, P_t \ .$$

A development of periodic autoregressions, similar to ordinary autoregressions, can be given (see Pagano (1978)) but we shall only briefly sketch these here. Given a sample $Y(1), \ldots, Y(T)$ where for simplicity we take $T = Nd$ with N a natural number, define the sample covariance kernel (ignoring the mean function, assumed zero)

$$R_N(k, v) = \frac{1}{N} \sum_{j=0}^{m} Y(k + dj)\, Y(v + dj)$$

$$k = 1, \ldots, d$$

$$v = 0, \ldots, T - k - 1$$

$$m = N - \left[\frac{k + v}{d}\right]$$

then $R_N(k, v) \xrightarrow[N \to \infty]{} R(k, v)$ a. s. .

Estimators of the a can be obtained from the normal equations

$$R_N(k, k - v) + \sum_{j=1}^{P_k} a_k(j)\, R_N(k - j, k - v) = \delta_{v,0}\, \hat{\sigma}_k^2, \quad v \geq 0$$

and if we define

$$\underset{\sim}{a}_k = \left(a_k(1), \ldots, a_k(P_k)\right)^T$$

$$\underset{\sim}{\hat{a}}_k = \left(\hat{a}_k(1), \ldots, \hat{a}_k(P_k)\right)^T$$

$$\Gamma_k = \left(R(k - j, k - \ell)_{\ j, \ell}\right)$$

one can show that, as in the Mann-Wald theorem, the $\hat{\sigma}_k^2$ are asymptotically independent and independent of the $\underset{\sim}{\hat{a}}_k$ and

$$L\left(\sqrt{N}\,(\underset{\sim}{\hat{a}}_k - \underset{\sim}{a}_k)\right) \underset{N \to \infty}{} N(\underset{\sim}{0},\ \hat{\sigma}_k^2\, \Gamma_k^{-1}) \qquad k = 1, \ldots, d$$

and the $\underset{\sim}{\hat{a}}_1, \ldots, \underset{\sim}{\hat{a}}_d$ are mutually independent.

Applying these ideas to the electricity data of Figure 9 we find the prediction models given in Figure 17. For example, the predictor for a Monday is the Monday average plus .7 times the deviation of

ELECTRICITY DATA

One Step Predictors

	P. Var	Var
Y(Mon) = $\overline{Y}$(Mon) + .7 Z (Sat)	11.0	21.8
Y(Tues) = $\overline{Y}$(Tues) + .56 Z(Sun)	3.6	14.5
Y(Wed) = $\overline{Y}$(Wed) + .79 Z(Tues)	6.0	16.0
Y(Thurs)= $\overline{Y}$(Thurs) + .56 Z(Wed) + .41 Z(Tues)	2.6	16.2
Y(Fri) = $\overline{Y}$(Fri) + .98 Z(Thurs)	0.6	15.3
Y(Sat) = $\overline{Y}$(Sat) + .95 Z(Fri)	1.1	11.3
Y(Sun) = $\overline{Y}$(Sun) + .94 Z(Sat)	1.0	9.2

Figure 17.

the previous Saturday from its average (Z) . The error of prediction (the prediction variances) are given in the PVAR column, and the last column is the daily variances about the daily means. It is interesting to note that the prediction formulas change from day to day, especially when comparing the beginning of the week to the end of the week. The daily variations are approximately equal, though Monday has the highest variability (21.8) , but when we look at the predictability of each day there are large differences. It is very hard to predict

Mondays whereas the end of the week (once industry has gotten into gear) is relatively easy to predict (see Parzen and Pagano (1977)).

One other advantage of periodic autoregressions is that they suggest an alternative parametrization for multiple autoregressions. If $\underset{\sim}{X}(\cdot)$ is a vector $(d \times 1)$ covariance stationary autoregressive process then it obeys the stochastic difference equation,

$$\underset{\sim}{X}(t) + \sum_{j=1}^{p} A(j)\, \underset{\sim}{X}(t-j) = \underset{\sim}{\varepsilon}(t)$$

where the $A(\cdot)$ are $d \times d$. If we define the mapping

$$X_j(t) = Y(j + d(t-1))$$

then $\underset{\sim}{X}(\cdot)$ is a multiple autoregression if and only if $Y(\cdot)$ is a periodic autoregression of period d . Capitalizing on the asymptotic independence of the $\underline{a}_k$ it is advantageous to treat multiple autoregressions as periodic autoregressions (see Pagano (1978)).

Acknowledgement

We thank Dr. H. Joseph Newton for helping with the plots, and Ms. Jessica Pollner for delivering the talk at Ames.

References

Akaike, H. (1972). Information theory and an extension of the maximum likelihood principle, in Proc. 2nd Int. Symp. Information Theory, Supp. to Problems of Control and Information Theory, 267-281.

Allais, M. (1961). Test de periodicité. Généralisation du test de Schuster au cas de séries temporelles autocorrelées dan l'hypothèse d'un processus de perturbations aléatoires d'un système stable. Bull. L'Inst. Internat. de Stat., Act. 33e Sess., XXXIX, 143-193.

Anderson, T. W. (1971). Statistical Analysis of Time Series. New York: Wiley.

Berk, K. (1974). Consistent autoregressive spectral estimates. Ann. Stat., 2, 489-502.

Boscovich, R. J. (1760). See, Boscovich and the combination of observations, by Churchill Eisenhart in Roger Joseph Boscovich, L. L. White (ed.), 1961, Fordham University Press, New York.

Carmichael, J.-P. (1978). Consistency of the autoregressive method of density estimation. Tech. Report No. 61, Statistical Science Division, SUNY at Buffalo.

Davis, H. T. and M. Pagano (1978). A modified Gram-Schmidt algorithm for fitting autoregressive schemes. To appear.

Gladyshev, E. G. (1961). Periodically correlated random sequences. Soviet Math. 2, 385-388.

Hannan, E. J. (1970). Multiple Time Series. New York: Wiley.

Jones, R. H. and W. M. Brelsford (1967). Time series with periodic structure. Biometrika, 54, 403-408.

Kromer, R. E. (1969). Asymptotic properties of the autoregressive spectral estimator. Tech. Report No. 13, Department of Statistics, Stanford.

Mann, H. B. and A. Wald (1943). On the statistical treatment of linear stochastic difference equations. Econometrica, 11, 173-220.

Masani, P. (1966). Recent trends in multivariate prediction theory. In Multivariate Analysis (P. R. Krishnaiah, ed.). New York: Academic Press, 351-382.

McClave, J. T. (1975). Subset autoregression. Technometrics, 17, 213-220.

Oberhofer, W. and J. Kmenta (1974). A general procedure for obtaining maximum likelihood estimates in generalized regression models. Econometrica, 42, 579-590.

Pagano, M. (1973). When is an autoregressive scheme stationary? Communications in Statistics, 1, 533-544.

Pagano, M. (1978). On periodic and multiple autoregressions. Annals of Statistics, 6.

Pagano, M. and M. J. Hartley (1975). On fitting distributed lag models subject to polynomial restrictions. Tech. Report No. 22, Statistical Science Division, SUNY at Buffalo.

Parzen, E. (1976). Multiple time series: determining the order of approximating autoregressive schemes. In Multivariate Analysis VI (P. Krishnaiah, ed.), New York: Academic Press.

Parzen, E. and M. Pagano (1977). An approach to modeling seasonally stationary time series. In Forecasting and Modeling Time-of-Day and Seasonal Electricity Demands (A. Lawrence, ed.), EA 578-SR. E. P. R. I., Palo Alto.