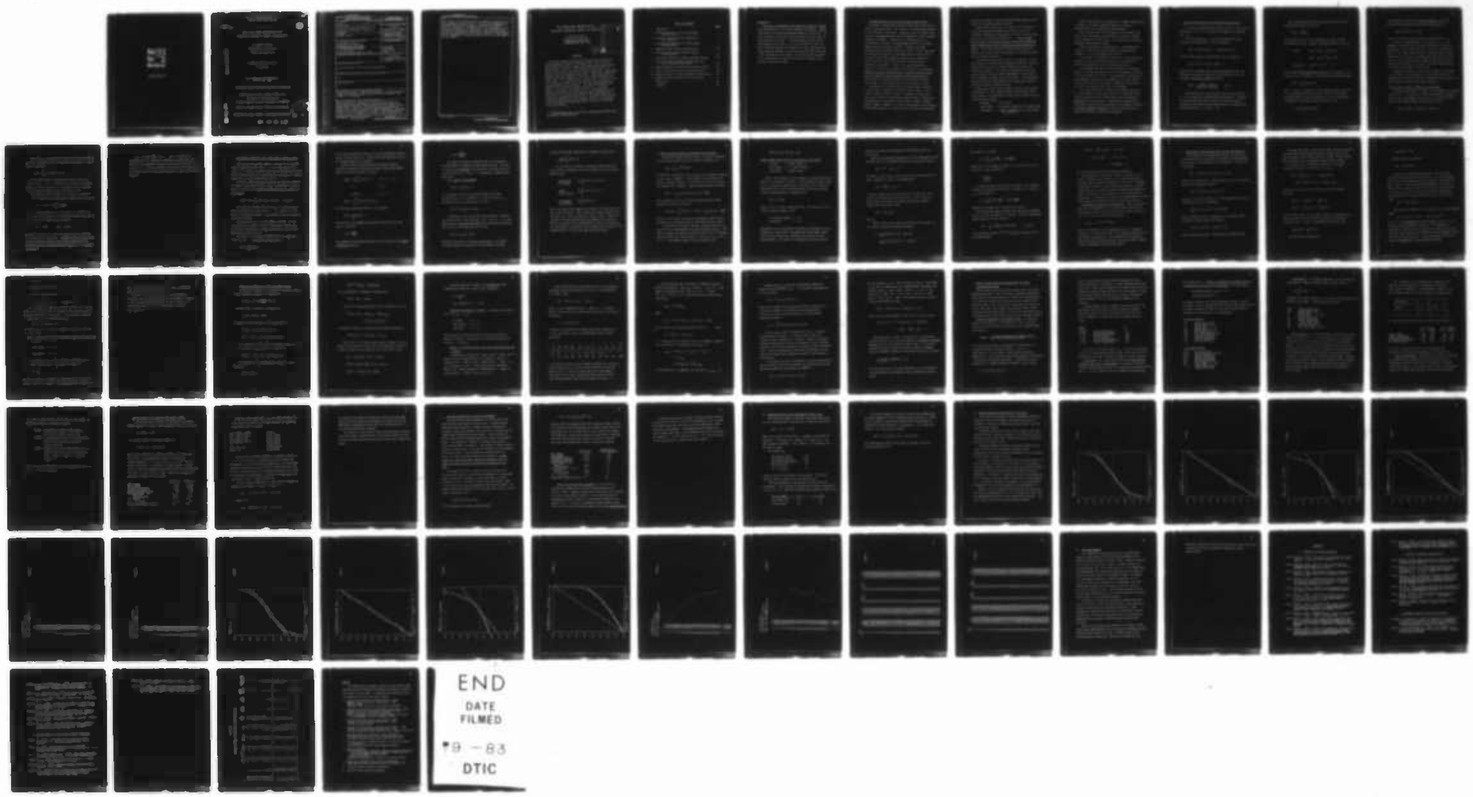
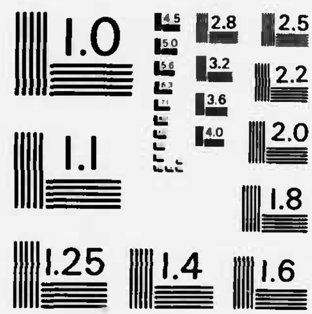
AD-A132 257 TIME SERIES MODEL IDENTIFICATION BY ESTIMATING INFORMATION MEMORY AND QUANTILES(U) TEXAS A AND M UNIV COLLEGE STATION DEPT OF STATISTICS E PARZEN JUL 83 1/1

UNCLASSIFIED TR-A-22 DAAG29-83-K-0051 F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143

STATISTICS
Phone 409 - 845-3141

(12)

TIME SERIES MODEL IDENTIFICATION BY ESTIMATING INFORMATION, MEMORY, AND QUANTILES

by Emanuel Parzen
Department of Statistics
Texas A&M University

Technical Report No. A-22
July 1983

Texas A&M Research Foundation
Project No. 4858

"Functional Statistical Data Analysis and Modeling"

Sponsored by the Army Research Office and
the Office of Naval Research
Contract DAAG29-83-K-0051, ONR N00014-83-WRM3008

Professor Emanuel Parzen, Principal Investigator

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER A-22	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Time Series Model Identification by Estimating Information, Memory and Quantiles		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Emanuel Parzen		8. CONTRACT OR GRANT NUMBER(s) DAAG29-83-K-0051 ONR N00014-83-WRM3008
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas A&M University Department of Statistics College Station, TX 77843		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE July 1983
		13. NUMBER OF PAGES 66
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

NA

18. SUPPLEMENTARY NOTES

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

Time series analysis, quantile data analysis, spectral analysis, ARMA representation, model identification, long memory time series, airline model.

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

This paper applies techniques of Quantile Data Analysis to non-parametrically analyze time series functions such as the sample spectral density, sample correlations, and sample partial correlations. The aim is to identify the memory type of an observed time series, and thus to identify parametric time domain models that fit an observed time series. Time series models are usually tested for adequacy by testing if their residuals are

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)

20.

white noise. It is proposed that an additional criterion of fit for a parametric model is that it have the non-parametrically estimated memory characteristics. An important diagnostic of memory is the index δ of regular variation of a spectral density; estimators are proposed for δ. Interpretations of the new quantile criteria are developed through cataloging their values for representative time series. The model identification procedures proposed are illustrated by analysis of long memory series simulated by Granger and Joyeux, and the airline model of Box and Jenkins.

TIME SERIES MODEL IDENTIFICATION BY ESTIMATING INFORMATION, MEMORY, AND QUANTILES

by Emanuel Parzen

Department of Statistics
Texas A&M University
College Station, TX 77843

Abstract

This paper applies techniques of Quantile Data Analysis to non-parametrically analyze time series functions such as the sample spectral density, sample correlations, and sample partial correlations. The aim is to identify the memory type of an observed time series, and thus to identify parametric time domain models that fit an observed time series. Time series models are usually tested for adequacy by testing if their residuals are white noise. It is proposed that an additional criterion of fit for a parametric model is that it have the non-parametrically estimated memory characteristics. An important diagnostic of memory is the index δ of regular variation of a spectral density; estimators are proposed for δ. Interpretations of the new quantile criteria are developed through cataloging their values for representative time series. The model identification procedures proposed are illustrated by analysis of long memory series simulated by Granger and Joyeux, and the airline model of Box and Jenkins.

Research supported in part by the Army Research Office and the Office of Naval Research.

Table of Contents

Dedication

This paper is dedicated to the memory of Gwilym M. Jenkins. The contributions to time series analysis of Gwilym M. Jenkins (1932-1982) will always be embedded deeply into the field. His work (especially joint work with George Box) has influence in diverse fields of science. I was fortunate to come to know Gwilym early in my career, on a visit to London in 1958. He spent 1959-1960 with me at Stanford and I spent 1961-1962 with him at Imperial College. He earned the respect and affection of all who knew him or his work. His life and work was heroic. As we contemplate the sadness of his death so young, may we continue to enjoy his spirit.

1. FUN.STAT approach to time series model identification

The need to analyze data arising in the form of time series arises in diverse fields. The concept of a conventional analysis is not the same in each field. Engineers tend to estimate mean, variance, and spectrum (which may be regarded as a non-parametric signature of models). Economists and forecasters tend to estimate mean, variance, and time domain models such as ARMA or ARIMA (which are parametric models). Spectral and ARMA estimation are not routine procedures; there are many algorithms for spectral estimation and time domain model identification. In addition there are critics of spectral and correlation based methods of time series analysis, of whom the most prominent is Mandelbrot (1982). This paper describes an approach to time series analysis which attempts to use diverse methods of analysis simultaneously in order to meet the needs of all the fields of applications of time series analysis. It also aims to integrate spectral and correlation methods with methods for long memory and/or long tailed time series.

An approach to spectral analysis and time domain modeling of time series is described in Parzen (1979), (1980), (1981), (1982), (1983a), (1983b), (1983c). An approach (motivated by time series methods) to statistical data analysis of probability distributions is described in Parzen (1979), (1982), (1983a), 1983b), (1983c), (1983d); it is called the Quantile Data Analysis and FUN.STAT approach, to connote that it is based on functional statistical inference, entropy and information measures, and quantile and density quantile approach.

Parzen (1980) states that "a criterion that any general time series modeling strategy must fulfill is that its conceptual framework should provide a role for the continuing quest for a time series decomposition. ... Thus it seems critical that a successful approach to time series modeling employ simultaneously both the spectral domain and the time domain." This paper discusses the enhanced insight to be obtained by also employing simultaneously the quantile domain and the information domain.

This paper discusses how to add to our approach to time series model identification new diagnostic measures, based on quantile data analysis of spectral density function, and information measures. The approach implemented in our time series computer program library TIMESBOARD is called ARSPID (for autoregressive spectral identification). The "enhanced" approach could be called ARSPIQ (for autoregressive spectral information quantile identification).

In empirical time series analysis a central role in model identification is the concept of memory [see Parzen (1981)] which yields a classification of a time series into one of the following three classes:

no memory	≡	white noise
short memory	≡	stationary ergodic but not white noise
long memory	≡	trends, seasonal cycles, long cycles, non-stationary .

When a time series is classified as no memory (white noise), it requires no further analysis (except for quantile identification of its probability distribution).

When a time series is classified as a short memory time series, it is described (parametrised) by ARMA(p,q) schemes that transform it to white noise. The orders p and q are not measures of the length of memory.

When a time series is classified as a long memory time series it is described (parametrised) by operators which transform it to a short memory time series.

To describe the dependence structure of a time series one introduces quantitative indices which are non-parametric statistics guiding our choice of parametric models.

An ARMA model (which is a finite parameter time domain model) is a parametric description of the dependence structure of a short memory time series. A nonparametric description of its dependence structure is provided by the spectral density function from which one can deduce "significant frequencies" (at which the spectral density has local maxima).

The operations which transform a long memory time series to a short memory one (or which represent a long memory time series in terms of a short memory one) can be considered a parametric time domain model. Nonparametric descriptions of long memory properties are introduced in this paper in terms of the index of regular variation of the spectral density at a specified frequency, usually zero frequency.

2. Quantile identification of probability distributions

To identify probability distribution that fit a time series sample $Y(t)$, $t=1,\ldots,T$, one treats the sample as a data batch $X_1,\ldots,X_n$.

For a data batch $X_1,\ldots,X_n$ one can define the sample distribution function $\tilde{F}(x)$, $-\infty<x<\infty$, defined by

$$\tilde{F}(x) = \text{fraction of } X_1,\ldots,X_n \text{ which are } \leq x,$$

and the sample quantile function $\tilde{Q}(u)$, $0\leq u\leq 1$, defined by

$$\tilde{Q}(u) = \tilde{F}^{-1}(u) = \inf\ \{x\colon F(x) \geq u\} \quad .$$

Quick and dirty insight into the distributions that fit the univariate distribution function $\tilde{F}$ is provided by a plot of the sample informative quantile function

$$I\tilde{Q}(u) = \frac{\tilde{Q}(u) - \tilde{Q}(0.5)}{2\{\tilde{Q}(0.75) - \tilde{Q}(0.25)\}} \quad , \ 0\leq u\leq 1 \quad .$$

The IQ function is plotted with a vertical scale from -1 to 1; its values are truncated when they exceed ± 1. For ease of interpretation of the IQ function, we also plot the IQ function of the uniform distribution which is a straight line passing through (0, -.5) and (1, .5).

The distribution functions $F(x)$ that we seek to fit to the data are usually of the form

$$F(x) = F_o(\frac{x-\mu}{\sigma})$$

for parameters μ and σ to be estimated, and $F_o(x)$ a known distribution function. The most important cases of $F_o(x)$ are:

normal $\quad F_o(x) = \Phi(x) = \int_{-\infty}^{x} \phi(y)\, dy \quad ,$

$$\phi(y) = (2\pi)^{-1/2} \exp - \frac{1}{2} y^2 \quad ;$$

exponential $\quad F_o(x) = 1 - e^{-x} \quad , \quad x > 0 \quad .$

One can <u>test</u> (<u>before parameter estimation</u>) the goodness of fit of $\tilde{F}(x)$ to $F(x) = F_o(\frac{x-\mu}{\sigma})$ by introducing the weighted spacings

$$\tilde{d}(u) = \frac{1}{\tilde{\sigma}_o} f_oQ_o(u)\, \tilde{q}(u)$$

where: $f_oQ_o(u) = f_o(F_o^{-1}(u))$ is the density-quantile function of the specified distribution; $\tilde{q}(u) = \tilde{Q}'(u)$ is the sample quantile density function (expressible in terms of spacings, or differences of successive order statistics); and

$$\tilde{\sigma}_o = \int_0^1 f_oQ_o(u)\, \tilde{q}(u)\, du$$

is an estimator of σ called the <u>score deviation</u>. The test function is the cumulative weighted spacings function

$$\tilde{D}(u) = \int_0^u \tilde{d}(t)\, dt, \quad 0 \leq u \leq 1$$

which one compares with the uniform distribution $D(u) = u$.

To test for exponentiality, take $f_0 Q_0(u) = 1 - u$. The diagnostic function $\tilde{D}(u)$ will appear linear when the data is exponential. In the important case of a mixture distribution, [that is, the lower order statistics represent values from an exponentially distributed sub-population], $\tilde{D}(u)$ will be linear over an initial interval $0 \leq u \leq p$. When the data batch is the sample spectral density, the value p estimates the proportion of the total power which is white noise.

Diagnostic measures of time series parameters [the sample spectral density and correlogram] are provided by plots of suitable $\tilde{IQ}(u)$ and $\tilde{D}(u)$ functions. Examples of their power as discriminators of memory are given in Section 7.

<u>Quantile Data Analysis of Sample Spectral Density</u>

When the sample mean $\bar{Y}$ is large, it is necessary to transform $Y(t)$ to $Y(t) - \bar{Y}$; otherwise one would always obtain a diagnostic that $Y(\cdot)$ is a long memory time series. An alternative first step in time series analysis is to replace $Y(t)$ by

$$\{Y(t) - \tilde{Q}(0.5)\} \div 2\{\tilde{Q}(0.75) - \tilde{Q}(0.25)\} \quad .$$

When Y(t) is a pre-processed time series (from the sample, the mean or median has been subtracted) one computes the sample Fourier transform

$$\tilde{\psi}(\omega) = \sum_{t=1}^{T} Y(t) \exp(-2\pi i \omega t)$$

at an equi-spaced grid of frequencies in $0 \leq \omega \leq 1$ of the form $\omega = k/S$, $k=0,1,\ldots,S-1$. We call S the spectral computation number; one should choose $S \geq T + M$, where M is the maximum lag at which one computes sample correlations $\tilde{\rho}(v)$.

The sample spectral density $\tilde{f}(\omega)$, $0 \leq \omega \leq 1$, is computed at $\omega = k/S$ by squaring and normalizing the sample Fourier transform:

$$\tilde{f}(\omega) = |\tilde{\psi}(\omega)|^2 \div \frac{1}{S} \sum_{k=0}^{S-1} |\tilde{\psi}(\tfrac{k}{S})|^2 \quad .$$

The classification of the time series as no memory (or white noise) is equivalent to the random variables representing the values of the sample spectral density

$$\tilde{f}(\omega), \quad \omega=k/S \qquad k=1,\ldots,[S/2] \quad ,$$

having the property that they are asymptotically independent and exponentially distributed. Therefore tests for white noise can be obtained by quantile data analysis based tests for exponentiality of the sample spectral density $\tilde{f}(\omega)$ at suitable frequencies.

The data batch $\tilde{f}(\frac{k}{S})$, k=0, 1,...,S/2, is tested for exponentiality by forming its informative quantile function $\tilde{IQ}(u)$ and its cumulative weighted spacings function $\tilde{D}(u)$, with $f_oQ_o(u) = 1-u$. How one interprets the quantile data analysis of the sample spectral density (periodogram) is best illustrated by examples.

3. Correlation diagnostics for model memory identification

The time series analyst seeks to develop for an observed sample time series $Y(t)$, $t=1,2,\ldots,T$ of a time series $Y(t)$, $t=0, \pm 1, \ldots$ various functions that can be estimated and plotted which provide insight into, and diagnostic measures of, possible models that fit the observed time series.

Schuster (1898) pioneered techniques of spectral analysis. To detect hidden periodicities, Schuster proposed calculating what we today call the sample unnormalized spectral density or periodogram

$$f_T(\omega) = \frac{1}{T} \left| \sum_{t=1}^{T} Y(t) \exp(-2\pi i t\omega) \right|^2 , \quad -0.5 \leq \omega \leq 0.5 .$$

One actually computes and plots $f_T(\omega)$ at an equi-spaced grid of frequencies $\omega_k = k/S$, $k=0,1,\ldots, S-1$, where S is the spectral computation number. Using the Fast Fourier Transform, one chooses $T < S \leq 2T$.

The graph of $f_T(\omega)$ is a very wiggly function. If one interprets local maxima of $f_T(\omega)$ as indicating "significant frequencies" representing "hidden periodicities" one obtains many spurious periodicities.

The notion of the spectral density $f(\omega)$ of a time series $Y(t)$, $t=0, \pm 1, \ldots$ is defined heuristically by

$$f(\omega) = \lim_{T \to \infty} f_T(\omega) \qquad .$$

If the limit existed one might call $f(\omega)$ the asymptotic spectral density of the time series. However the limit does not exist in any customary mode of convergence.

Wiener (1930) proposed solving the harmonic analysis problem by defining the sample covariance function $R_T(v)$ which equals the Fourier transform of $f_T(\omega)$:

$$
\begin{aligned}
R_T(v) &= \frac{1}{T}\sum_{t=1}^{T-v} Y(t+v)\,Y(t) \quad &, \quad & v = 0, 1, \ldots, T-1 \\
&= 0 &, \quad & v \geq T, \\
&= R_T(-v) &, \quad & v < 0 \quad ;
\end{aligned}
$$

$$R_T(v) = \int_{-0.5}^{0.5} \exp(2\pi i v\omega)\, f_T(\omega) \quad .$$

The limit whose existence needs to be assumed is

$$R(v) = \lim_{T\to\infty} R_T(v) \quad ;$$

one calls $R(v)$ the asymptotic covariance function of the time series. One calls

$$\rho(v) = \frac{R(v)}{R(0)}$$

the asymptotic correlation function; it is the limit of the sample correlation function

$$\rho_T(v) = \frac{R_T(v)}{R_T(0)} \quad .$$

The sample correlation function $\rho_T(v)$ is an important building block for methods of model identification. Its plot is called the correlogram. One could test for white noise by testing whether $\rho_T(v)$, $v=1,2,..,N$ constitute a random normal data batch.

The cumulative periodogram

$$F_T(\omega) = \int_0^{\omega} f_T(\omega')\, d\omega$$

is a diagnostic tool for providing evidence of hidden periodicities. If it converges, its limit function $F(\omega)$ provides a spectral representation of $R(v)$:

$$R(v) = \int_0^1 \exp 2\pi i v\omega \, dF(\omega) \quad .$$

A probability model under which the asymptotic covariance functions exists is the following: $Y(t)$, $t=0, \pm 1, \ldots$ is a zero mean Gaussian covariance stationary time series with covariance function $R(v)$ satisfying (for all t and v)

$$R(v) = E[Y(t+v)\, Y(t)] \quad .$$

When the time series is stationary and ergodic, the sample covariance function converges to the covariance function.

A Gaussian stationary time series is ergodic if and only if

$$\lim_{T\to\infty} \frac{1}{T} \sum_{v=1}^{T} R^2(v) = 0 \quad .$$

It is natural to classify a stationary time series into three classes according to the rate of decay of the correlation function $\rho(v)$:

<u>white noise</u> (no memory)	$\frac{1}{T} \sum_{v=1}^{T} \rho^2(v) = 0$ for all T
<u>ergodic</u> (short memory)	$\frac{1}{T} \sum_{v=1}^{T} \rho^2(v) \to 0$ as $T \to \infty$
<u>non-ergodic</u> (long memory)	$\frac{1}{T} \sum_{v=1}^{T} \rho^2(v) \not\to 0$.

One of the aims of this paper is to discuss the unifying role of the concept of memory. The foregoing trichotomy indicates that there are three types of memory (no, short, long). However the insights into model identification provided by the notion of memory are captured not by definitions in terms of correlations (or even partial correlations) but by definitions in terms of the spectral density function and sample spectral density.

4. Spectral density memory classification and indices

The spectral density function $f(\omega)$, $-0.5 \le \omega \le 0.5$ is defined as the Fourier transform of the correlation function $\rho(v)$:

$$f(\omega) = \sum_{v=-\infty}^{\infty} e^{-2\pi i v\omega} \rho(v) \quad .$$

A sufficient condition for $f(\omega)$ to exist as an ordinary function is that $\rho(v)$ is summable. A long memory time series may not possess a spectral density. To be able to use such a function, we introduce the sequence of approximating spectral densities

$$\bar{f}_T(\omega) = \sum_{|v|<T} \exp(-2\pi i v\omega)\, \rho(v)\, (1 - \frac{|v|}{T}) \quad .$$

The correlation criteria for memory classification provide equivalent criteria in terms of

$$\mathrm{Var}\,[\bar{f}_T] = \int_{-0.5}^{0.5} \{\bar{f}_T(\omega) - 1\}^2 \, d\omega = 2 \sum_{v=1}^{T} \rho(v)\, (1 - \frac{|v|}{T})^2 \quad .$$

However a more useful criterion is the dynamic range of $\bar{f}_T(\omega)$. We discuss its definition only for the case that $f(\omega)$ exists.

A stationary time series can have a spectral density $f(\omega)$ and yet not be representable as an autoregressive process. One needs to assume an additional condition such as $f(\omega)$ is bounded above and below; for some constants c_1 and c_2, $0 < c_1 \le f(\omega) \le c_2 < \infty$. The dynamic range of $f(\omega)$ is defined to be

$$\{\max_{\omega} \log f(\omega) - \min_{\omega} \log f(\omega)\} .$$

Dynamic range classification of memory of a time series:

no memory	$\equiv$	dynamic range = 0
short memory	$\equiv$	$0 <$ dynamic range $< \infty$
long memory	$\equiv$	dynamic range $= \infty$.

Often, zero frequency is the frequency at which the spectral density has a behavior causing it to have infinite dynamic range. As $\omega \to 0$, the spectral density $f(\omega)$ is assumed to be a regularly varying function, with the representation [called the regular variation representation at frequency $\omega=0$]

$$f(\omega) = \omega^{-\delta} L(\omega)$$

where $L(\omega)$ is a slowly varying function. The value of δ is an index of length of memory, since

No and short memory	$\equiv$	$\delta = 0$
Long memory	$\equiv$	$\delta \neq 0$.

Long memory time series models considered by Mandelbrodt (1973), Granger and Joyeux (1980), and Geweke and Porter-Hudak (1983) have spectral density $f(\omega)$ satisfying the regular variation representation. The index $\delta<0$ corresponds to a zero value for

$f(\omega)$ at $\omega=0$, while $\delta>0$ corresponds to an infinite value for $f(\omega)$ at $\omega=0$.

When $\delta>0$, the spectral density $f(\omega)$ is an integrable function only for $0\leq\delta<1$; the correlation function $\rho(v)$ decays slowly as

$$\rho(v) \sim v^{\delta-1} \quad \text{as } v \to \infty \quad .$$

The value at $\omega=0$ of $f(\omega)$ can be ∞ and still $\delta=0$; this holds for $f(\omega) \sim (\log\omega)^2$ for small ω, corresponding to

$$\rho(v) \sim \frac{\log v}{v} \quad \text{as } v \to \infty \ .$$

A symbolic spectral density $f(\omega)$ with $\delta>1$ is that of a time series $Y(\cdot)$ whose first difference $\Delta Y(t) = Y(t) - Y(t-1)$ is short memory (covariance stationary with spectral density bounded above and below); then

$$f_Y(\omega) \sim \frac{1}{\omega^2} f_{\Delta Y}(\omega)$$

and $\delta=2$.

Parzen (1983d) gives explicit formulas for the index δ in the context of density-quantile estimation:

$$\begin{aligned}\delta &= \lim_{\omega\to 0} \int_0^1 \log f(\omega y)\, dy - \log f(\omega) \\ &= \lim_{\omega\to 0} \frac{1}{\omega} \int_0^{\omega} \log f(\lambda)\, d\lambda - \log f(\omega) \quad .\end{aligned}$$

To estimate δ one forms

$$\delta_k = \frac{1}{k} \sum_{j=1}^{k} \log f(\frac{j}{n}) - \log f(\frac{k+1}{n})$$

where n and k are integers tending to ∞ in such a way that k/n tends to 0. One can show that

$$\delta = \lim_{\substack{k\to\infty \\ k/n\to 0}} \delta_k \quad .$$

A similar formula can be used to estimate δ in a regular variation representation of $f(\omega)$ at a frequency ω_o: represent $\omega_o = m/n$ and define

$$\delta_k = \frac{1}{k} \sum_{j=1}^{k} \log f(\frac{j+m}{n}) - \log f(\frac{k+1+m}{n}) \quad .$$

Examples of estimates of δ are given in Section 7.

We estimate the memory index δ from consistent estimators $\hat{f}(\omega)$ of the spectral density f. We use: (1) the non-parametric kernel spectral density estimator

$$\hat{f}(\omega) = \sum_{v=-\infty}^{\infty} k(\frac{v}{M}) \rho_T(v) \exp -2\pi i \omega v \quad , \quad |\omega| \leq 0.5$$

with truncation point $M = T^{7/8}$ (in practice, we use M = T/2) and Parzen window

$$k(t) = 1 - 6t^2 + 6|t|^3, \quad |t| \le 0.5,$$

$$= 2\,(1-|t|)^3 \quad , \quad 0.5 \le |t| \le 1,$$

$$0 \quad , \quad \text{otherwise} \quad ;$$

and (2) autoregressive spectral density estimators.

Only examples can show which values of δ occur in real series. The goal in estimating δ is to develop diagnostics concerning the "detrending" operations to be used to transform a long memory series to a short memory time series. To model time series, Box and Jenkins (1970) introduced the ARIMA(p,d,q) model. Estimation of the parameter d can be approached by estimating δ. Estimation of p and q can be approached by diverse order determining methods involving estimating information.

<u>Determining the degree of differencing</u>: When a time series Y(t) can be transformed to a stationary time series Z(t) by differencing d times, one can think of the "spectral density" $f_Y(\omega)$ of $Y(\cdot)$ as having the representation

$$f_Y(\omega) = |1-e^{-2\pi i\omega}|^{-2d} f_Z(\omega)$$

which is a special case of assuming that $f_Y(\omega)$ is regularly varying at $\omega=0$ with index $\delta=2d$. The foregoing estimators for δ may provide alternatives to the techniques for estimating d which have been proposed by Granger and Joyeux (1980), Janacek (1982), and Geweke and Porter-Hudak (1983).

5. ARMA models and prediction error memory classification

The concept of an autoregressive process was introduced by Yule (1927) as an alternative technique for detecing hidden periodicities, and estimation of the frequency ω in the time series model

$$Y(t) = A \cos 2\pi\omega t + B \sin 2\pi\omega t + \varepsilon(t)$$

where $\varepsilon(\cdot)$ is white noise. The function $\cos 2\pi\omega t$ satisfies the second order difference equation

$$Y(t) + a_1 Y(t-1) + a_2 Y(t-2) = 0$$

with $a_1 = -2 \cos 2\pi\omega$ and $a_2 = 1$. Yule suggested determining coefficients a_1 and a_2 minimizing

$$\sum_{t=1}^{T} \{Y(t) + a_1 Y(t-1) + a_2 Y(t-w)\}^2 \quad .$$

These coefficients may be interpreted as estimators of the parameters in the "random shock" model

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

where $\varepsilon(t)$ is white noise. Thus was born the AR(2) model.

Autoregressive (AR), moving average (MA), and autoregressive-moving average schemes (ARMA) now play a central role in time series analysis, since they provide basic models for time series model identification, forecasting, and spectral estimation.

One definition of an ARMA(p,q) model for a zero mean covariance stationary time series Y(t), t=0, ± 1, ... is

$$Y(t) + a_p(1)\, Y(t-1) + \ldots + a_p(p)\, Y(t-p)$$

$$= \varepsilon(t) + b_q(1)\, \varepsilon(t-1) + \ldots + b_q(q)\, \varepsilon(t-q)$$

where $\varepsilon(t)$ is a white noise time series, and the transfer functions

$$g_p(z) = 1 + a_p(1)z + \ldots + a_p(p)\, z^p,$$

$$h_q(z) = 1 + b_q(1)\, z + \ldots + b_q(q)\, z^q$$

have all their roots in the complex z-plane in the region $|z|>1$. For the backward shift operator B we use the lag operator L, defined by LY(t) = Y(t=1). An ARMA(p,q) model is written

$$g_p(L)\, Y(t) = h_q(L)\, \varepsilon(t) \quad .$$

An AR(∞) model is expressed

$$g_\infty(L)\ Y(t) = \varepsilon(t) \quad .$$

An MA(∞) model is expressed

$$Y(t) = h_\infty(L)\ \varepsilon(t) \quad .$$

A model for a stationary time series is an invertible filter which transforms it to white noise. For a short memory time series, the whitening filters can always be represented as AR(∞) or MA(∞) and are approximated by ARMA(p,q) of suitable orders to be estimated. The white noise $\varepsilon(t)$ to which we seek to transform a time series $Y(t)$ are the infinite memory one step ahead prediction errors (innovations) $Y^\nu(t) = Y(t)-Y^\mu(t)$, where

$$Y^\mu(t) = E[Y(t)|Y(t-1),\ldots\ ] \quad .$$

The white noise sequence $Y^\nu(t)$ has mean 0 and variance $\sigma_\infty^2 R(0)$, where

$$\sigma_\infty^2 = E[|Y^\nu(t)|^2] \div R(0), \quad R(0) = E[|Y(t)|^2] \quad .$$

We call σ_∞^2 the <u>normalized</u> mean square prediction error, of one-step ahead infinite memory prediction. The importance of normalization (which may not currently be standard practice for all time series analysts) is emphasized by the information theory approach in the next section. A basic diagnostic tool is the memory m <u>normalized</u> mean square prediction errors

$$\sigma_m^2 = E[|Y^{\nu,m}(t)|^2] \div R(0),$$

$$Y^{\nu,m}(t) = Y(t) - Y^{\mu,m}(t)$$

$$-Y^{\mu,m}(t) = a_m(1)\ Y(t-1) + \ldots + a_m(m)\ Y(t-m) \quad .$$

Given a true (or sample) correlation function $\rho(v)$, one can compute (using the Yule-Walker equations) the sequence σ_m^2 which converges monotonely to the limit σ_∞^2. An alternative approach to computing σ_∞^2 is the fundamental formula

$$\log \sigma_\infty^2 = \int_0^1 \log f(\omega)\ d\omega \ .$$

The value of σ_∞^2 is a very useful diagnostic measure of the memory of a time series.

<u>Memory classification by Normalized Mean Square Prediction Error</u>

no memory $\equiv$ $\sigma_\infty^2 = 0$

short memory $\equiv$ $0 < \sigma_\infty^2 < \infty$

long memory $\equiv$ $\sigma_\infty^2 = 1.$

The estimation of σ_∞^2 is one of the basic problems of time series model identification. One important method is

$$\hat{\sigma}_\infty^2 = \hat{\sigma}_{\hat{m}}^2$$

where $\hat{m}$ is chosen by an order-determining criterion (AIC due to Akaike or CAT due to Parzen). The pioneering work of Akaike (1974),

(1977) has shown the central role of information theoretic ideas in defining these criteria.

The next section discusses how to use information divergence ideas to measure the ability of ARMA(p,q) schemes to provide approximating models to the exact models (of a short memory time series) provided by AR(∞) and MA(∞) representations.

6. Information approach to memory and ARMA schemes

Information divergence of a probability density g from a (true) probability density f is defined by

$$I(f;g) = \int_{-\infty}^{\infty} \{-\log \frac{g(y)}{f(y)}\} f(y)\, dy \quad .$$

Information has an important decomposition

$$I(f;g) = H(f;g) - H(f)$$

defining cross-entropy H(f;g) and entropy H(f) by

$$H(f;g) = \int_{-\infty}^{\infty} \{-\log g(y)\} f(y)\, dy \quad ,$$

$$H(f) = H(f;f) = \int_{-\infty}^{\infty} \{-\log f(y)\} f(y)\, dy \quad .$$

The information $I(Y|X)$ about a continuous random variable Y in a continuous random vector X is defined by

$$I(Y|X) = I(f_{Y|X};\, f_Y) = E_X\, I(f_{Y|X};\, f_Y) \quad .$$

The entropy of Y and conditional entropy of Y given X are defined by

$$H(Y) = H(f_Y)$$

$$H(Y|X) = H(f_{Y|X}) = E_X H(f_{Y|X=x}) \quad .$$

One can establish a fundamental decomposition

$$I(Y|X) = H(Y) - H(Y|X) \quad .$$

Define the information about Y in X_2 conditioned on X_1 by

$$I(Y|X_1;\ X_1,X_2) = H(f_{Y|X_1}) - H(f_{Y|X_1,X_2})$$

$$= H(Y|X_1) - H(Y|X_1,X_2) \quad .$$

A fundamental formula to evaluate an information increment is

$$I(Y|X_1;X_1,X_2) = I(Y|X_1,X_2) - I(Y|X_1) \quad .$$

When X and Y are jointly normal random variables, let $\Sigma(Y)$ denote the variance of Y and $\Sigma(Y|X)$ the conditional variance of Y given X (which does not depend on the value of X). Then

$$H(Y) = \frac{1}{2} \log \Sigma(Y) + \frac{1}{2} (1 + \log 2\pi)$$

$$H(Y|X) = \frac{1}{2} \log \Sigma(Y|X) + \frac{1}{2} (1 + \log 2\pi)$$

$$I(Y|X) = - \frac{1}{2} \log \Sigma^{-1}(Y)\, \Sigma(Y|X) \quad .$$

A general approach to memory uses <u>information</u> in the infinite past about the current value, defined by

$$I_\infty = \lim_{m\to\infty} I_m \quad .$$

$$I_m = I(Y(m+1)\,|\,Y(1),\ldots,Y(m)) \quad .$$

<u>Information Definition of Memory</u>. We define a time series $Y(t)$, $t=0, \pm 1, \ldots$ to be

$$\text{no memory} \equiv I_\infty = 0 \quad ;$$
$$\text{short memory} \equiv 0 < I_\infty < \infty \quad ;$$
$$\text{long memory} \equiv I_\infty = \infty \quad .$$

This definition agrees with the criterion in the previous section in terms of σ_∞^2 since for a stationary Gaussian time series $I_\infty = -\frac{1}{2}\log \sigma_\infty^2$.

<u>Example. A random walk has long memory and white noise has no memory</u>.

A random walk is defined by $Y(m+1) = Y(m) + \varepsilon(m+1)$, $Y(0) = 0$, where $\varepsilon(t)$ are independent $N(0,\sigma^2)$, $\Sigma(Y(m+1)) = (m+1)\,\sigma^2$, $E[Y(m+1)\,|\,Y(1),\ldots,Y(m)] = Y(m)$, $\Sigma(Y(m+1)\,|\,Y(1)1,\ldots,Y(m)) = \sigma^2$, $I_m = \frac{1}{2}\log(m+1)$, $I_\infty = \infty$. A pure white noise is defined by $Y(m) = \varepsilon(m)$. Then $\Sigma(Y(m+1)) = \sigma^2$, $E[Y(m+1)\,|\,Y(1),\ldots,Y(m)] = 0$, $\Sigma(Y(m+1)\,|\,Y(1),\ldots,Y(m)) = \sigma^2$, $I_m = 0$, $I_\infty = 0$.

Both a random walk and a pure white noise can be regarded as special cases [corresponding to $\rho=1$ and $\rho=0$ respectively] of the AR(1) model

$$Y(t) = \rho Y(t-1) + \varepsilon(t), \quad t=1,2,\ldots$$

where $\varepsilon(t)$ are independent $N(0,\sigma^2)$. When $|\rho| < 1$, an AR(1) defines a stationary (or asymptotically stationary) time series satisfying

$$I_\infty = -\frac{1}{2}\log(1-\rho^2).$$

In order to transform one's thinking about AR(1) models from ρ to I_∞ one needs a table of corresponding values of these parameters.

ρ	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95
I_∞	.005	.020	.047	.087	.144	.223	.337	.511	.830	1.16
I_∞	.25	.5	.75	1.0	1.25	1.50	1.75	2	3	4
ρ	.627	.795	.881	.930	.958	.975	.985	.991	.999	.9998

A very quick and dirty rule for memory diagnosis is to regard an observed value of $I_\infty \geq 1.5$ as an early detector of very long memory, and $I_\infty \geq 1.00$ as an early detector of long memory. This rule is to be used in conjunction with other rules for discriminating memory type which are given in Section 7.

We next discuss how to interpret an ARMA(p,q) scheme in terms of information. Let $I_{p,q} = I(Y|Y_{-1},\ldots,Y_{-p}, Y^{\nu}_{-1},\ldots Y^{\nu}_{-q})$ denote the information about Y(t) in Y(t-1),...,Y(t-p), $Y^{\nu}(t-1),\ldots,Y^{\nu}(t-q)$. For a Gaussian stationary short memory time series

$$I_{p,q} = -\frac{1}{2}\log \sigma^2_{p,q}$$

where

$$\sigma^2_{p,q} = \Sigma^{-1}(Y)\ \Sigma(Y|Y_{-1},\ldots,Y_{-p}, Y^{\nu}_{-1},\ldots,Y^{\nu}_{-q}).$$

Let Y^- denote the infinite past Y(t-1), Y(t-2), ... Then

$$I_\infty = I(Y|Y^-) = -\frac{1}{2}\log \sigma^2_\infty \quad .$$

A measure of the goodness of fit of an ARMA(p,q) model to the true model for a stationary time series is

$$\begin{aligned} I_{p,q;\sigma} &= I(Y|Y_{-1},\ldots,Y_{-p}, Y^{\nu}_{-1},\ldots,Y^{\nu}_{-q};\ Y^-) \\ &= I_\infty - I_{p,q} \\ &= (-\frac{1}{2}\log \sigma^2_\infty) + \frac{1}{2}\log \sigma^2_{p,q} \quad . \end{aligned}$$

A time series Y(·) is ARMA(p,q) if, and only if, $I_{p,q;\infty} = 0$.

Formulas for $I_{p,q;\infty}$ are most conveniently developed in terms of the coefficients $\beta_1, \beta_2, \ldots$ of the MA(∞) representation of a time series:

$$Y(t) = Y^{\nu}(t) + \beta_1 Y^{\nu}(t-1) + \ldots \quad .$$

There are two methods for estimating the MA(∞) coefficients; invert AR(m) where m is chosen by an order-determining criterion, or derive β_k from estimators of (the cepstral pseudo-correlations)

$$\psi(v) = \int_{-0.5}^{0.5} \exp(2\pi i v\omega) \log f(\omega)\, d\omega \quad .$$

In the Gaussian case, information is (up to a constant) the logarithm of variance. It may seem that there is no reason to prefer information to variance. However information concepts are meaningful even for non-Gaussian series (although they have not yet been extensively calculated in the non-Gaussian case). Thus by translating variance into information, one can eventually transfer one's Gaussian intuition to non-Gaussian data analysis.

To illustrate the use of information in model identification, let us consider the loss one sustains in using the best fitting AR(2) model when the true model is an ARMA(1,1)

$$Y(t) + a\, Y(t-1) = \varepsilon(t) + b\, \varepsilon(t-1) \quad .$$

One can compute σ_∞^2, $\rho(1)$, $\rho(2)$ in terms of a and b. The values of $\rho(1)$ and $\rho(2)$ determine (via the Yule-Walker equations) the optimal values $\hat{\sigma}_2$, $\hat{a}_2(1)$, $\hat{a}_2(2)$. When a = -.5, b = .5, one obtains $\sigma_\infty^2 = .4286$, $\rho(1) = .7143$, $\rho(2) = .3571$; $\hat{\sigma}_2^2 = .4418$, $\hat{a}_2(1) = -.9378$, $\hat{a}_2(2) = .3126$. The information loss in using the approximating AR(2) model

$$Y(t) - .9378\, Y(t-1) + .3126\, Y(t-2) = \varepsilon(t)$$

rather than the exact ARMA(1,1) with -a=b=.5 is .015, since

$$I(Y|Y_{-1},Y_{-2};\, Y^-) = \{-\tfrac{1}{2}\log \sigma_\infty^2\} - \{-\tfrac{1}{2}\log \hat{\sigma}_2^2\}$$

$$= .4236 - .4084 = .015 \quad .$$

Estimating MA(∞) is also a prerequisite to using another criterion that we use to estimate memory: the Prediction Variance Horizon function, introduced in Parzen (1981). It provides a quantitative method of measuring memory (especially medium memory) by HORIZON, defined as the smallest value of h for which

$$\frac{1 + \beta^2(1) + \ldots + \beta^2(h-1)}{1 + \beta^2(1) + \ldots} \geq 0.95 \quad .$$

The left hand side of the above inequality can be interpreted as representing the mean square error of prediction h steps ahead.

7. Quantile based time series diagnostics, and their representative values

This section introduces various quantile based time series diagnostic measures. Their use can be considered exploratory data analysis since they require no theory for interpretation if one is willing to base one's conclusions on the empirically observed values of the criteria for representative time series. On the other hand, the criteria are based on clearly stated concepts of probability theory, and one could study theoretially the distribution of the criteria for various time series models.

Quantile diagnostics of normality of data. A diagnostic measure of the shape of a distribution is the log standard deviation of the informative quantile function, denoted LNSDIQ, and defined by

$$\text{LNSDIQ} = \log\left\{\frac{\text{standard deviation of original data}}{\text{twice interquartile range}}\right\} .$$

For a normal distribution, interquartile range equals 1.35 standard deviation; therefore LNSDIQ = - log 2.7 = -1 approximately. We can regard a significant difference of LNSDIQ from -1 as an indication that the probability distribution of the data is not normal (Gaussian). A more formal test of normality is to compare LNSDIQ with LNSGMO = log $\tilde{\sigma}_0$, where

$$\tilde{\sigma}_0 = \int_0^1 \Phi^{-1}(u)\ \tilde{IQ}(u)\ du .$$

is the score deviation (an efficient estimator of σ for a normal distribution, obtained as a linear combination of order statistics). This test (analogous to the Shapiro-Wilk test for normality) requires further theory as we find examples in which the data have IQ(u) plots that are not normal (confirmed by LNSDIQ different from -1), yet LNSDIQ and LNSGMO are not different.

To decide whether data is normal, the entire graph of the informative quantile [IQ(u)] function should be examined. However an early detector of the shape is provided by the value of LNSDIQ as is indicated by the following empirical values:

LNSDIQ		$\hat{I}_\infty$
Variable	Cauchy white noise	0
-1.14	Airlines log monthly	1.38
-1.14	NYC Monthly Births	.93
-1.24	Lines + Noise	1.72
-1.34	Cauchy random walk	1.48
-1.34	NYC Monthly Temperature	1.17
-1.32	Normal random walk	1.11

In the tables in this section, $I_\infty = -\frac{1}{2}\log \sigma_\infty^2$ is estimated by $I_{\hat{m}}$ for the approximating AR(m) scheme, where the order $\hat{m}$ is determined by the AIC criterion (or equally the CAT criterion).

Periodogram. For a white noise time series whose random variables have finite second moment, the quantile function of the periodogram should be that of an exponential distribution with mean 1. A test of white noise is provided by examining IQ(u) for exponentiality. Powerful discriminators of memory type are the median and variance of the periodogram. For white noise

Periodogram median = log 2 = .69 ,

Periodogram variance = 1.

As memory increases, per. median decreases and per. variance increases, as the following empirical results confirm [the values for AR(1) processes are based on the table "Quantile Memory Analysis of Simulated AR(1)" in the Appendix].

Periodogram median

.89	Cauchy white noise
.7	Normal white noise
.2	Normal AR(1), ρ = .8
.08	Normal AR(1), ρ = .9
.02	Normal AR(1), ρ = .99
.08	NYC Births Monthly
.06	NYC Temperatures Monthly
.04	Normal random walk
.03	Airlines log monthly
.03	Cauchy random wlak
.02	Lines plus noise

Periodogram variance

67.7	Lines plus noise
49.8	NYC Temperatures Monthly
41.5	Normal random walk
38.3	Cauchy random walk
39.7	Airline log monthly
33.1	NYC Births monthly
42.	Normal AR(1), ρ = .99
22.	Normal AR(1), ρ = .9
1	Normal white noise
.5	Cauchy white noise

Correlations. As a memory diagnostic, we use correlations mean square of sample correlation $\hat{\rho}(v) = \rho_T(v)$, v=1,2,...,

$$\frac{1}{N} \sum_{v=1}^{N} \hat{\rho}^2(v) \quad ,$$

computed for a large value of N. It is zero for white noise, and increases with memory. Some empirical values are:

.002	Cauchy white noise
.004	Normal white noise
.01	Normal AR(1), $\rho = .7$
.1	Normal AR(1), $\rho = .9$
.2	Normal AR(2), $\rho = .99$
.14	NYC Births monthly
.18	Normal random wlak
.17	Cauchy random walk
.19	Airlines log monthly
.23	Line plus noise
.26	NYC Temperatures monthly

Delta estimators. A conclusion that a time series is long memory is regarded by us as valid only when it is confirmed by the behavior of the sequence of estimators δ_k of the memory index δ. We routinely form these estimators at ω=0 and ω=1/12. Note that 1/12 is the period of an annual cycle in monthly data; the program permits the specification of any other seasonal frequency. Two sequence of estimators δ_k are formed; from the best approximating AR scheme, and from Parzen window eatimators with truncation point approximately equal to T/2, where T is the time series sample size [the time series examined had T=144 to 200].

Our "estimator" $\hat{\delta}$ is currently only a summary of the behavior of the sequences δ_k, indicating a value about which there is clustering. For normal AR(1) schemes at $\omega=0$ the following typical values were found in simulated series.

approximate $\hat{\delta}$	2	1.5	1
when I_∞	1.75, 2	1.25, 1.50	1
ρ	.99	.96	.93

For empirical series we observed the following estimators $\hat{\delta}$:

	$\omega = 0$		$\omega = 1/12$	
	Best AR	Parzen window	Best AR	Parzen window
Lines + Noise	1.98	2.22	.38	.51
Cauchy random walk	1.84	1.84	.37	.48
Airlines log monthly	2.33	2.22	1.56	1.42
NYC Temperatures Monthly	-.4	-.8	2.1	2.6
NYC Births Monthly	2.05	1.74	1.12	.77

Note that a negative value of δ at $\omega=0$ indicates the possibility that the spectral density $f(\omega)$ is zero at $\omega=0$.

Partial correlations. The sequence of partial correlations are usually used to diagnose if the time series obeys an autoregressive scheme, since AR(p) is equivalent to partial correlations equal to 0 for orders greater than p. The quantile function of partial correlations then should look like white noise plus as many outliers as the order of the scheme. As diagnostic measures of memory we compute:

PCIQR = interquartile range of the quantile function of partial correlations;

PCLNSD = log standard deviation of the informative quantile function IQ(u) of partial autocorrelations;

PCOUT = number of partial correlations greater in absolute value than twice interquartile range, number of values of u at which $|IQ(u)| > 1$.

Typical values of these measures for representative time series will be published elsewhere.

8. ARSPIQ analysis of simulated long memory series

To illustrate their research on long memory time series models, Granger and Joyeux (1980) generated series of the form

$$(I-L)^d Y(t) = \varepsilon(t)$$

with spectral density (for some constant c)

$$f_Y(\omega) = c(1 - \cos 2\pi\omega)^{-d} \quad .$$

This spectral density is regularly varying at $\omega=0$ with memory index $\delta=2d$. They generated two series of length 400, corresponding to $d = .25$ ($\delta=.5$) and $d = .45$ ($\delta=.9$). We call these series White $\delta.5$ and White $\delta.9$ respectively. I would like to thank Clive Granger and Roselyne Joyeux for having given us copies of their series to study. Some of the diagnostics generated by ARSPIQ are as follows:

	White $\delta.5$	White $\delta.9$
DATA LNSDIQ	-.95	-1.03
DATA LNSGMO	-.95	-1.03
Variance Periodogram	6.9	10.9
Median Periodogram	.54	.30
Correlation Mean Square	.02	.03
Delta Estimator $\omega=0$		
Best AR	0.9	1.0
Parzen Window	0.6	1.2
AIC order $\hat{m}$	7	4
$I_\infty = -\frac{1}{2}\log\hat{\sigma}^2_{\hat{m}}$	.14	.35
Prediction Variance Horizon	24	20

Comparing these diagnostics with the values obtained for various series in Section 7, we might conclude the following characteristics for the series.

Data LNSDIQ, LNSGMO	Normal
Corr. Mean Square	Short memory
Periodogram, Var	Short memory
Periodogram, Median	Short memory
I_∞	Short memory
Pred. Var. Hor.	Medium memory
Delta $\omega=0$	Long memory

Printer plots of delta estimators are given in Figures 5, 6, 11, 12. One does not currently get an exact numerical estimate of δ. But the values estimated for δ are consistent wiht the theoretical values of δ used in generating the time series. On the basis of the foregoing diagnostics, one would be justified in recommending a fractional differencing of the time series, using a rough estimate of δ.

If one fitted an ARMA model to these series one might be tempted to fit ARMA(1,1) models: for white δ.5,

$$Y(t) - .75\ Y(t-1) = \varepsilon(t) - .47\ \varepsilon(t-1)\ ;$$

for white δ.9,

$$Y(t) - .89\ Y(t-1) = \varepsilon(t) - .44\ \varepsilon(t-1)\quad .$$

By comparing the spectral distribution function of these ARMA schemes with the cumulative periodogram one would see that the ARMA models inadequately modeled the low frequency portion of the spectral distribution function.

The question is open whether expect practictioners of purely time domain ARMA or ARIMA methods of time series analysis could identify the model generating the series simulated by Granger and Joyeux.

9. Does the airline data fit the airline model?

The aim of time series modeling is to find a filter that transforms the time series to white nosie. A possible model identification procedure is to guess a model, estimate its parameters, form the residuals, and test if the residuals are not significantly different from white noise. This procedure in practice may lead two different analysts to infer two different models. The question is open how to resolve which model to accept (which model is "better"). The concept of memory seems to provide a characteristic of a time series which can be estimated non-parametrically. Statisticians must decide whether to accept as a model fitting criterion the following: a model fitted to a time series must satisfy the criterion that its memory characteristics agree with those estimated from the data.

The operation of this criterion can be illustrated by a classic series used as a test case by researchers on time series model identification methods — log international airlines passengers series. The model fitted by Box and Jenkins (1970) to this series has become celebrated as the "airline model". It takes 1st and 12th differences of the series Y(t) to form a short memory time series $\tilde{Y}(t)$:

$$(I-L)(I-L^{12})\ Y(t) = \tilde{Y}(t);$$

$\tilde{Y}(\cdot)$ is modeled as a special form of MA(12):

$$\tilde{Y}(t) = (I-\theta_1 L)(I-\theta_{12} L^{12})\ \varepsilon(t).$$

Parzen (1982) has suggested that 12th differences might suffice as an operation which transforms the original series (which has long memory) to a new series which is just barely short memory. The diagnostics in the table [which one interprets by comparing them with the representative values in Section 7] indicate that 12th differencing does suffice to yield short memory.

	Log Airline	Log Airline 12th difference
Data LNSDIQ	-1.15	-.97
Data LNSGMO	-1.16	-.97
Periodogram Median	.03	.19
Periodogram Variance	39.7	7.7
Correlation Mean Sq.	.19	.05
Delta Estimate $\omega=0$		
Best AR	2.33	0
Parzen Window	2.22	0
Delta Estimate $\omega = 1/12$		
Best AR	1.56	0
Parzen Window	1.42	0
$\hat{I}_\infty = -\frac{1}{2}\log\hat{\sigma}^2_{\hat{m}}$	1.38	.5
Prediction variance horizon	51	66+

Note on how we form the estimator $\hat{\delta}$: we write $\hat{\delta}=0$ to indicate that sequence $\hat{\delta}_k$ oscillates between negative and positive values. Negative values could indicate $\hat{\delta}<0$ and presence of a zero of the spectral density. In our current state of knowledge we assign a value to $\hat{\delta}$ representing essentially flat behavior of δ_k. If the 12th difference spectral density had a zero at $\omega=0$ or $\omega/1/12$, we would suspect that we had over-differenced.

A quantitative measure of memory is the prediction variance horizon [51 for airline, >66 for 12th difference]; one concludes that differencing the time series still has significant trend components (long memory). The ARARMA modeling procedure of Parzen (1982) finds that if one transforms the airline series by the operator $I - 1.02L^{12}$ rather than by $I - L^{12}$, one does obtain a time series which is unequivocably short memory.

10. <u>ARSPIQ Analysis of 12th difference of white noise</u>

The ability of ARSPIQ to identify time series models may be well illustrated by an analysis of a simulated time series

$$Y(t) = \varepsilon(t) - \varepsilon(t-12),$$

where $\varepsilon(t)$ is N(0,1) white noise. A sample of size T=200 was simulated. It had mean .02, median .01, variance 2.16. The DATA diagnostics LNSDIQ = -1.04, LNSGMO = -1.04 indicate that the data is normal.

The diagnostics

Periodogram median	.38
Periodogram variance	2.63
Correlation mean square	.01
Best AR order $\hat{m}$	24
$\hat{I}_\infty = -\frac{1}{2} \log \hat{\sigma}^2_{\hat{m}}$	.27

indicate that the time series is short memory. But the AR spectral density estimator does not perform well.

The delta diagnostics indicate that the time series is long memory. That the spectral density has zeroes at frequencies $\omega=0$ and $\omega= 1/12$ is indicated by significantly negative values of δ:

Delta estimate	$\omega=0$	$\omega = 1/12$
Best AR($\hat{m}$ = 24)	-1.9	-1.2
Parzen window	-1.6	- .9

To estimate prediction variance horizon [and an ARMA scheme by select regression on the covariance matrix of Y(t-j), Y (t-k)] we fit an MA(∞) by inverting an AR(96) whose coefficients are computed by a Burg algorithm; it estimates $I_\infty = .63$, prediction horizon > 100, and chooses the model

$$Y(t) + .41\, Y(t-12) = \varepsilon(t) - .55\, \varepsilon(t-12).$$

This ARMA spectral density has exactly the shape of the true spectral density of Y(·).

11. Quantile graphics printer plots illustrated

The printer plot graphical output generated by ARSPIQ is illustrated for the long memory simulated series White δ.5 and White δ.9 which are respectively labelled JOY1 and JOY2 on the attached output.

Informative quantile function of the original time series JOY1 and JOY2 are plotted in Figure 1 and 7 respectively (with letters O and M); $\tilde{IQ}(u)$ plots indicate normality, confirmed by $\tilde{D}(u)$ plots in Figures 2 and 8.

Informative quantile function of the periodogram of time series JOY1 and JOY2 are plotted in Figures 3 and 9 respectively; they are not exactly exponential, as is confirmed by $\tilde{D}(u)$ plots in Figures 4 and 10.

The index δ of regular variation of the spectral density at zero frequency is estimated by the "limit" of the sequence δ_k plotted in Figures 5 and 11 (using AR spectral density estimator) and Figures 6 and 12 (using Parzen window spectral density estimator). In Figure 5, a limit exists which is approximately 0.9; in figure 6, one may assign a limit value of approximately 0.6. In figure 11 the limit is assigned to be approximately 1; in figure 12, the limit is assigned to be approximately 1.2.

Figures 13 and 14 represent covariances of the time series $Y(t)$ and its innovations $\varepsilon(t) = Y^{\nu}(t)$ estimated for input into the "ARMA identification by select regression" procedure. The last column is Prediction Variance Horizon function.

FIGURE 1

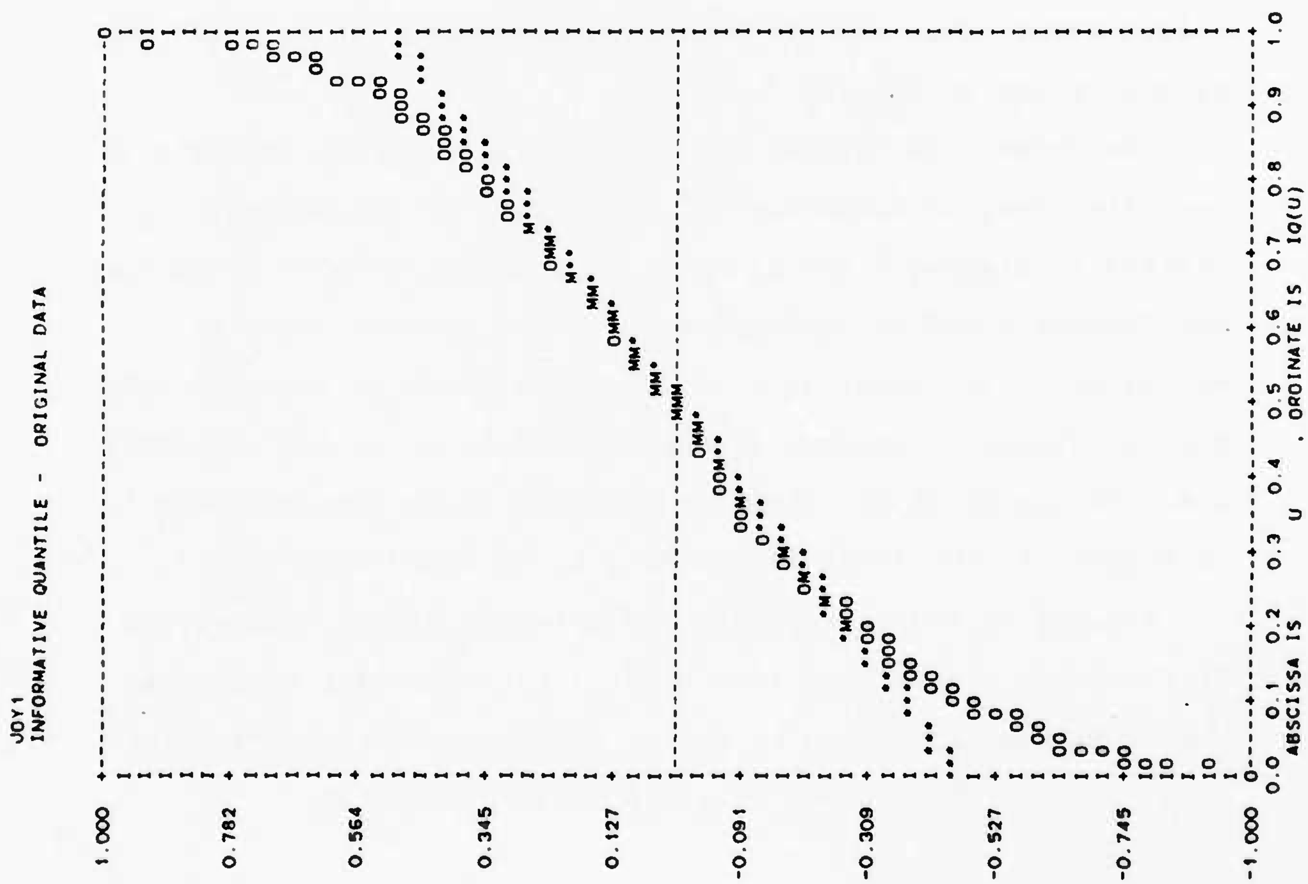

JOY1
INFORMATIVE QUANTILE - ORIGINAL DATA
1.000
0.782
0.564
0.345
0.127
-0.091
-0.309
-0.527
-0.745
-1.000
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
ABSCISSA IS U , ORDINATE IS IQ(U)

FIGURE 2

RAW DISTRIBUTION D(U) NORMAL CASE
D+ = 0.8572 AT U = 0.3137 . D- = -0.1677 AT U = 0.9862

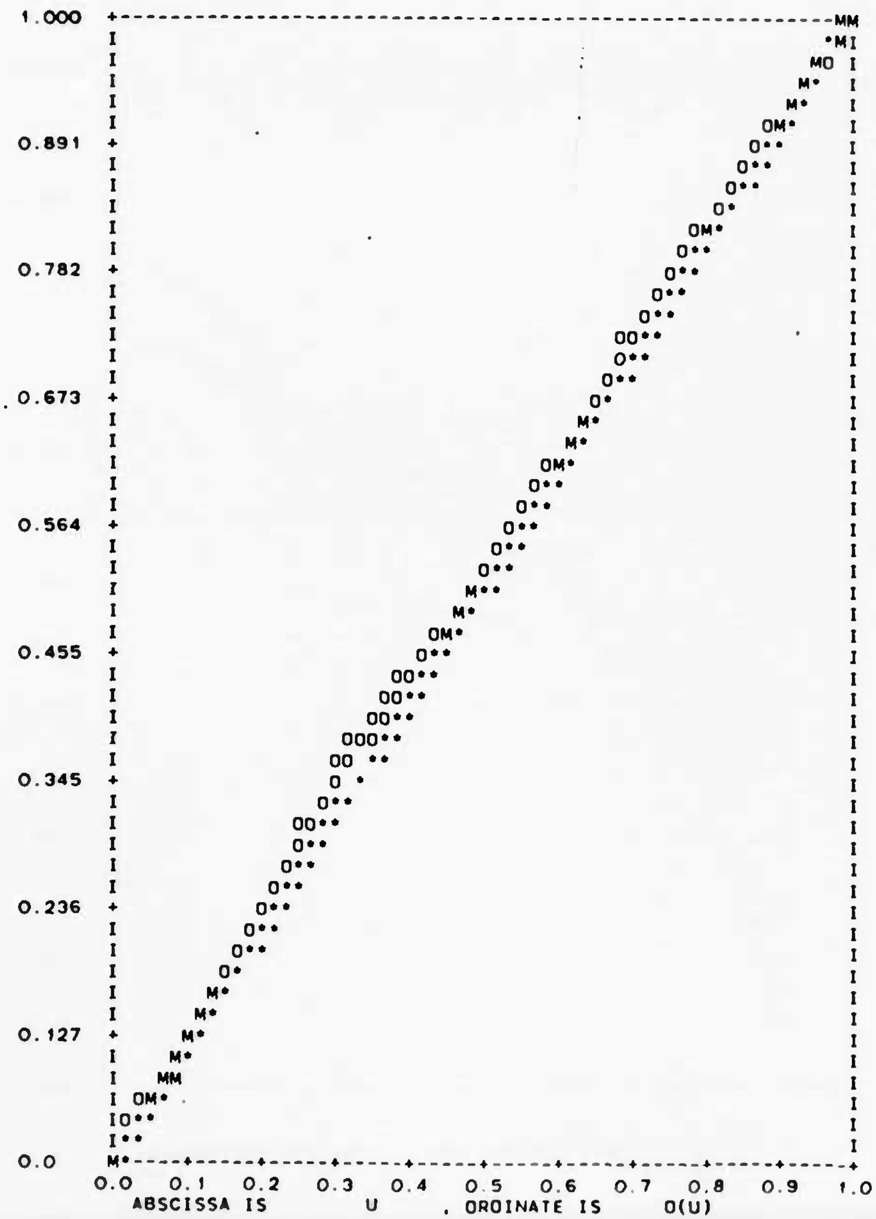
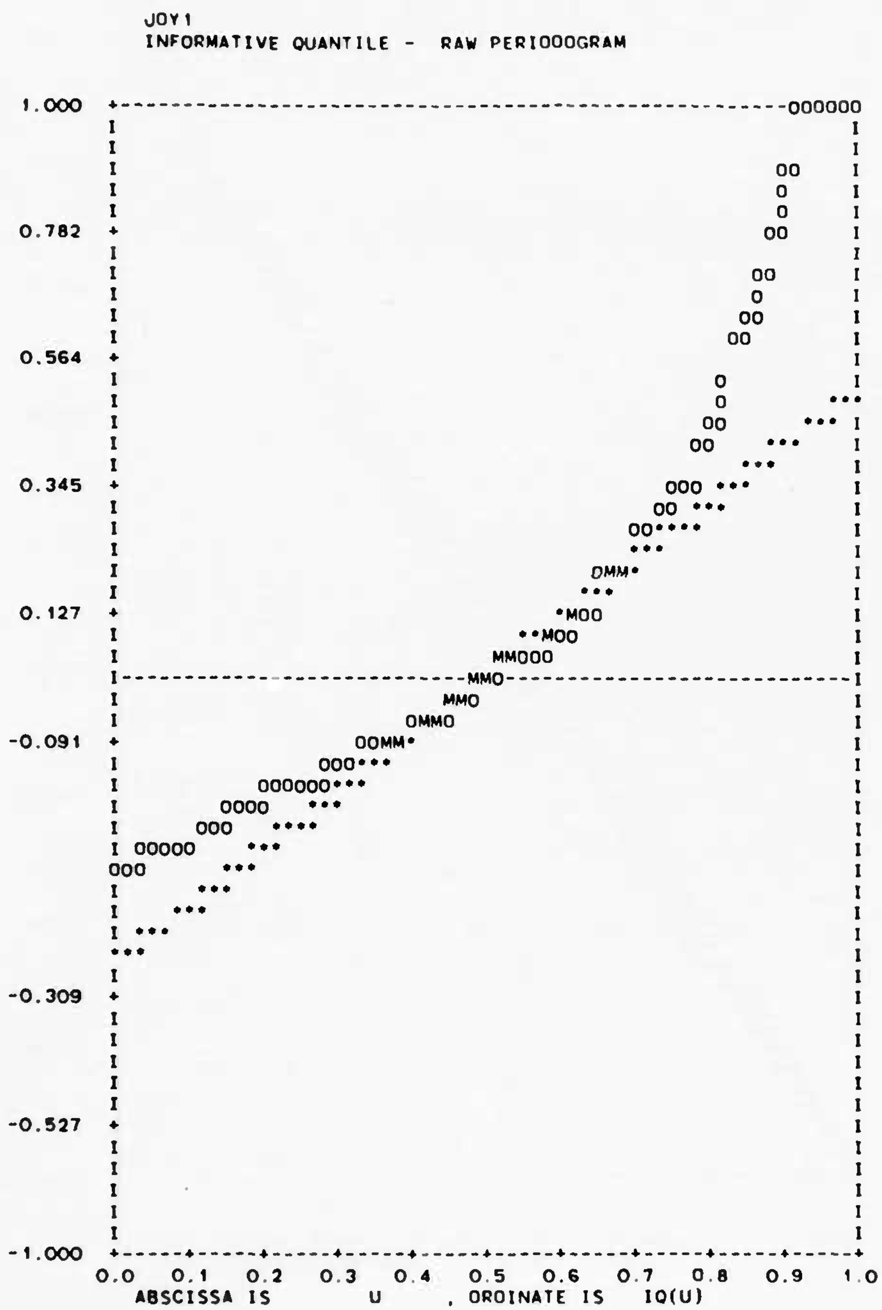


FIGURE 3



JOY1
INFORMATIVE QUANTILE - RAW PERIODOGRAM
1.000
0.782
0.564
0.345
0.127
-0.091
-0.309
-0.527
-1.000
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
ABSCISSA IS U , ORDINATE IS IQ(U)

RAW DISTRIBUTION D(U) EXPONTAL CASE
D+ = 0.0663 AT U = 0.9971 . D- = -4.4206 AT U = 0.9150

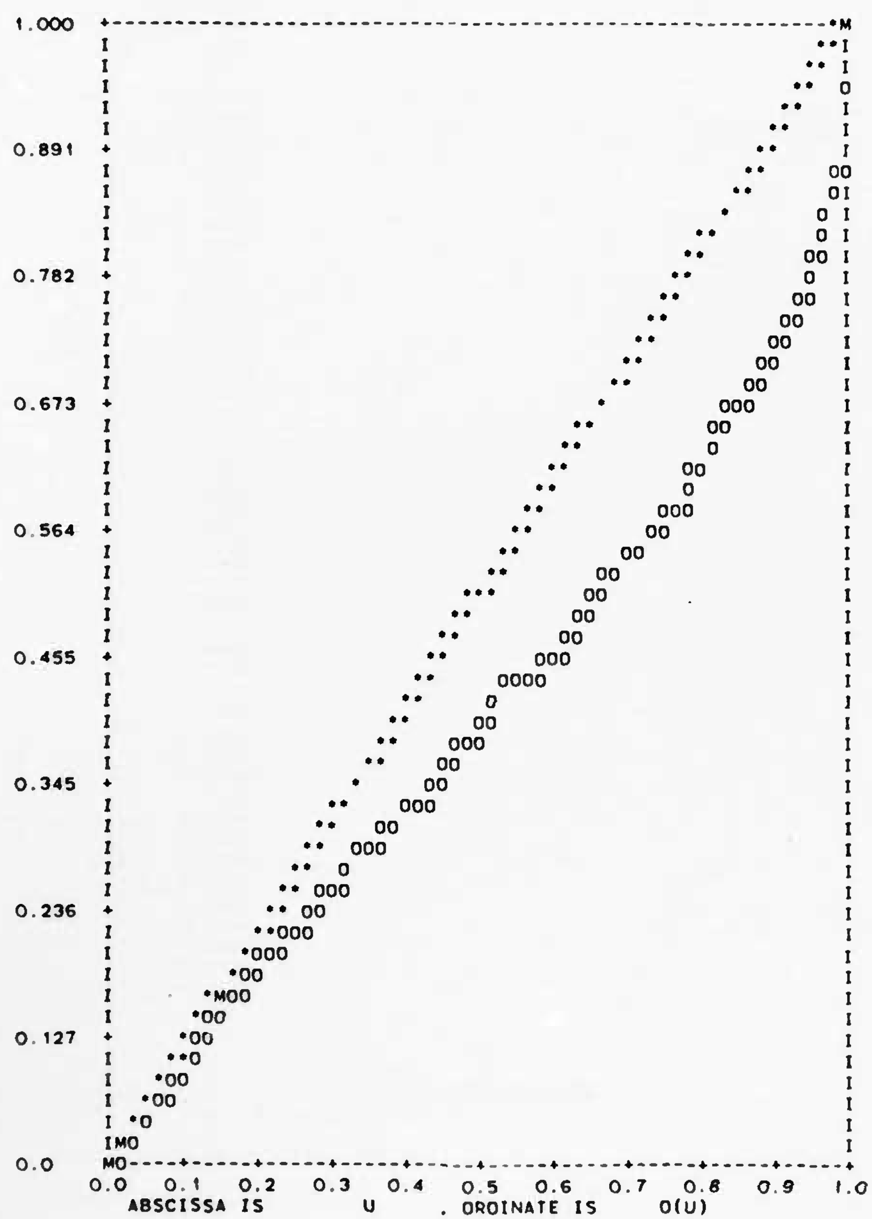


FIGURE 4

FIGURE 5

DELTA MEMORY FUNCTION

BEST ORDER AR SPECTRAL DENSITY

PLOT 1 - LAG 1 IS AT FREQUENCY O
PLOT 2 - LAG 1 IS AT FREQUENCY KSEAS/NFREQS

I	DELT
1	0.0174
2	0.0590
3	0.1176
4	0.1879
5	0.2645
6	0.3432
7	0.4207
8	0.4949
9	0.5646
10	0.6291
11	0.6882
12	0.7420
13	0.7906
14	0.8344
15	0.8735
16	0.9083
17	0.9392
18	0.9663
19	0.9900
20	1.0105
21	1.0281
22	1.0428
23	1.0550
24	1.0647
25	1.0722
26	1.0775
27	1.0807
28	1.0820
29	1.0815
30	1.0792
31	1.0752
32	1.0696
33	1.0625
34	1.0539
35	1.0438
36	1.0324
37	1.0196
38	1.0056
39	0.9903
40	0.9738
41	0.9561
42	0.9374

I	DELT
1	0.0020
2	0.0010
3	-0.0013
4	-0.0048
5	-0.0094

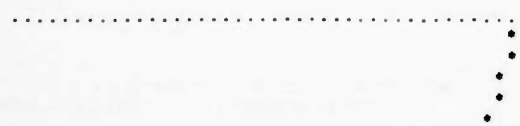


FIGURE 6

DELTA MEMORY FUNCTION

SMOOTHED PERIODOGRAM - PARZEN WINDOW

PLOT 1 - LAG 1 IS AT FREQUENCY 0
PLOT 2 - LAG 1 IS AT FREQUENCY KSEAS/NFREQS

I	DELT
1	0.0400
2	0.1385
3	0.2838
4	0.4654
5	0.6651
6	0.8547
7	1.0023
8	1.0871
9	1.1109
10	1.0920
11	1.0479
12	0.9879
13	0.9151
14	0.8322
15	0.7456
16	0.6667
17	0.6092
18	0.5847
19	0.5983
20	0.6435
21	0.7000
22	0.7373
23	0.7301
24	0.6784
25	0.6081
26	0.5521
27	0.5333
28	0.5608
29	0.6325
30	0.7387
31	0.8648
32	0.9937
33	1.1083
34	1.1943
35	1.2435
36	1.2551
37	1.2370
38	1.2036
39	1.1706
40	1.1504
41	1.1470
42	1.1536

I	DELT
1	0.0262
2	0.0097
3	-0.0409

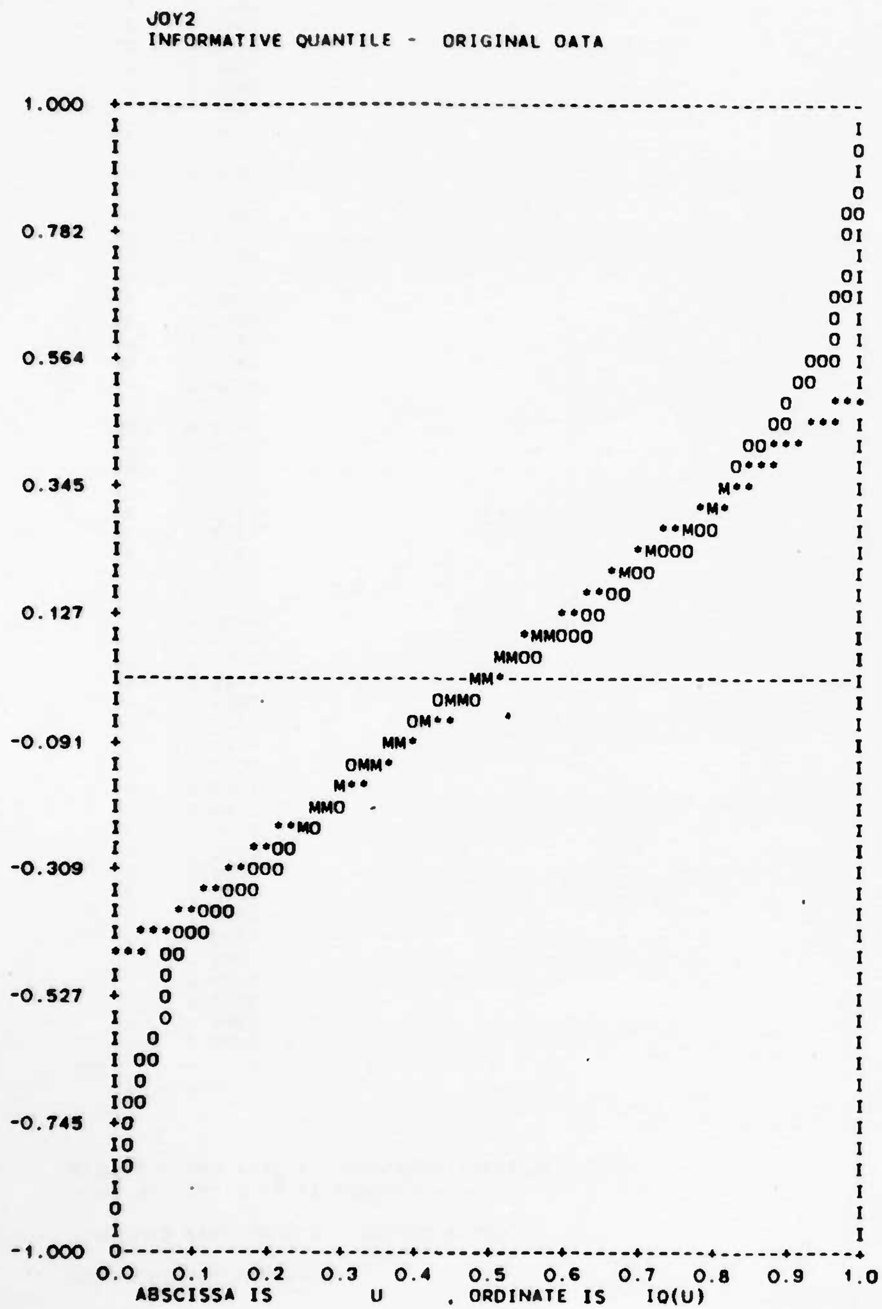
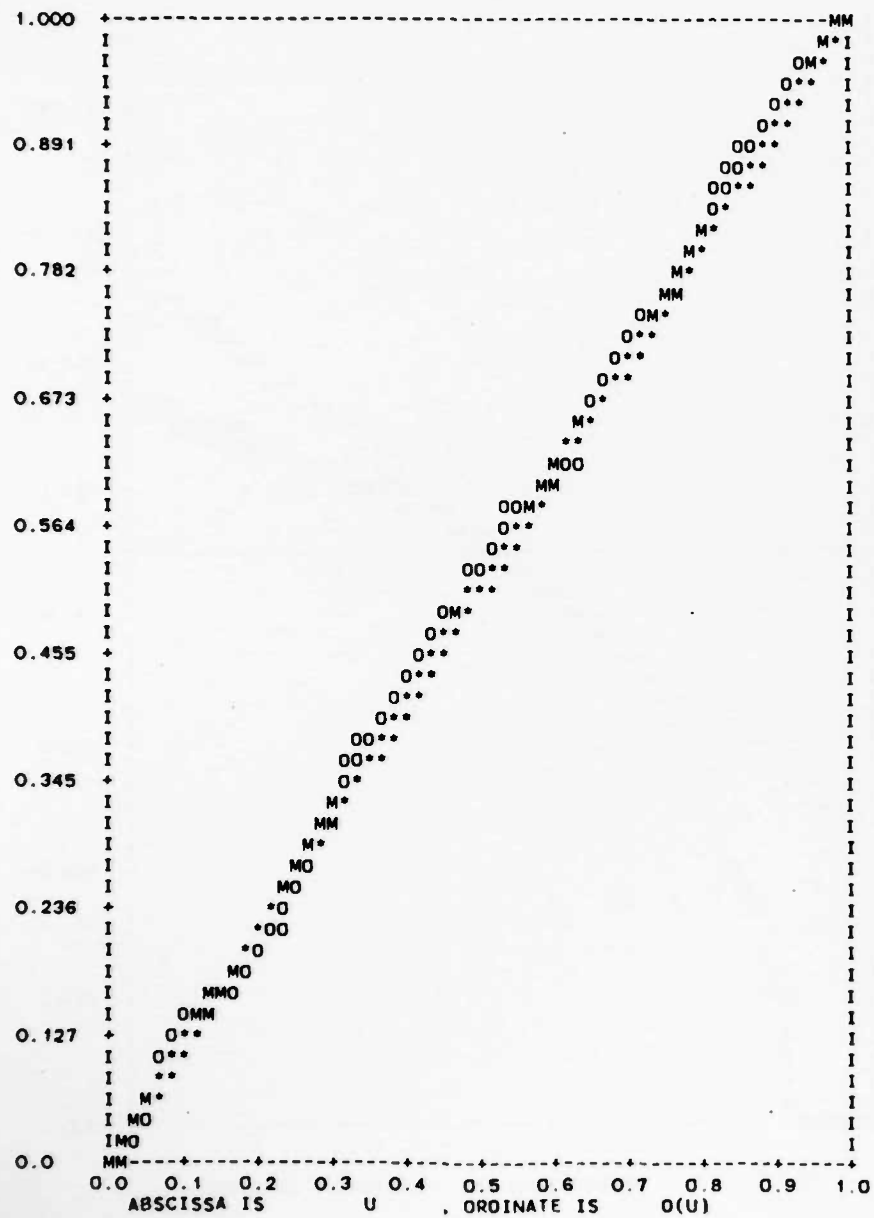

JOY2
INFORMATIVE QUANTILE - ORIGINAL DATA
1.000
0.782
0.564
0.345
0.127
-0.091
-0.309
-0.527
-0.745
-1.000
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
ABSCISSA IS U , ORDINATE IS IQ(U)

FIGURE 7

FIGURE 8

RAW DISTRIBUTION O(U) NORMAL CASE
O+ = 0.6127 AT U = 0.8462 . O- = -0.3379 AT U = 0.2162



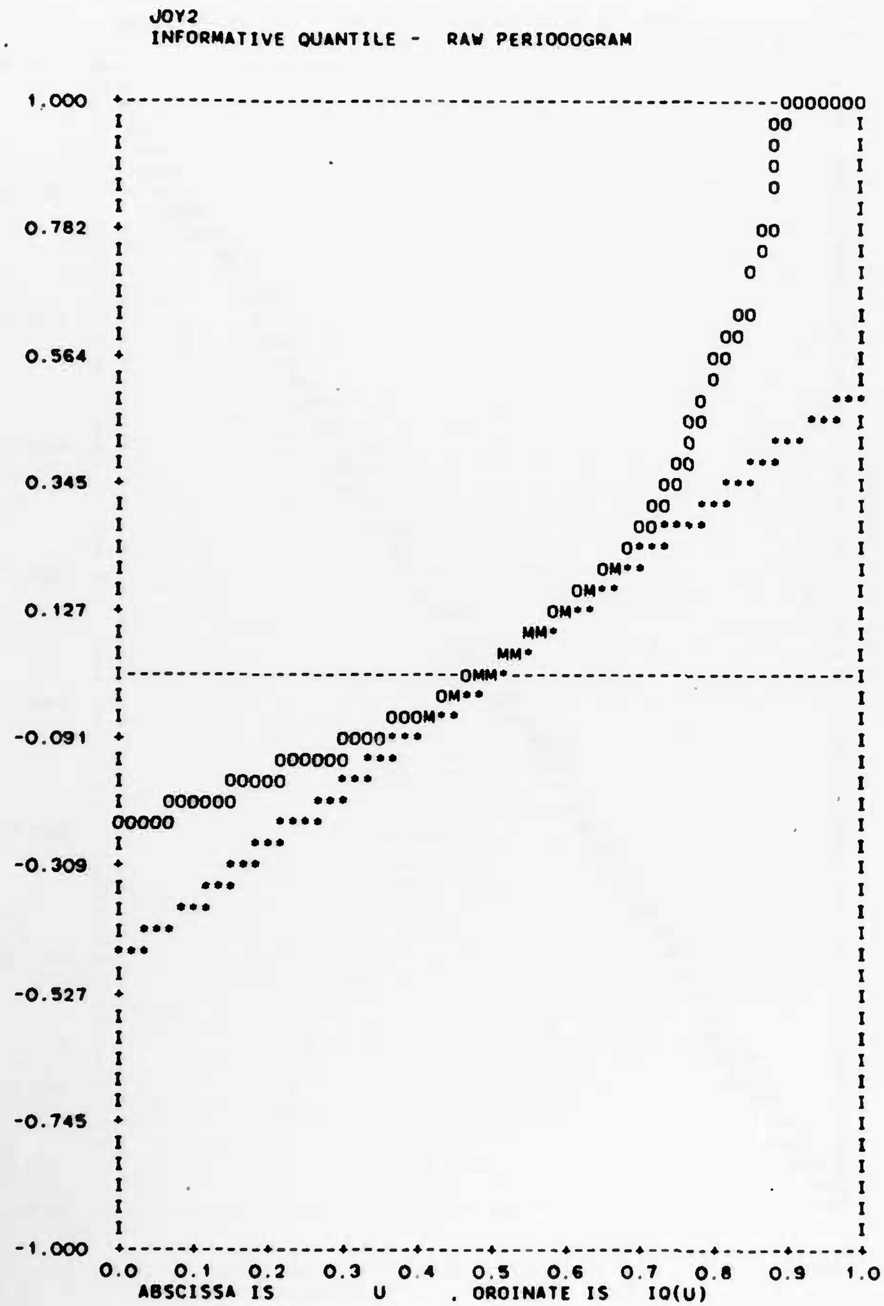

JOY2
INFORMATIVE QUANTILE - RAW PERIODOGRAM
1.000
0.782
0.564
0.345
0.127
-0.091
-0.309
-0.527
-0.745
-1.000
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
ABSCISSA IS U . ORDINATE IS IQ(U)

FIGURE 9

FIGURE 10

RAW DISTRIBUTION D(U) EXPONTAL CASE
D+ = 0.0978 AT U = 0.9932 . D- = -9.7469 AT U = 0.9385

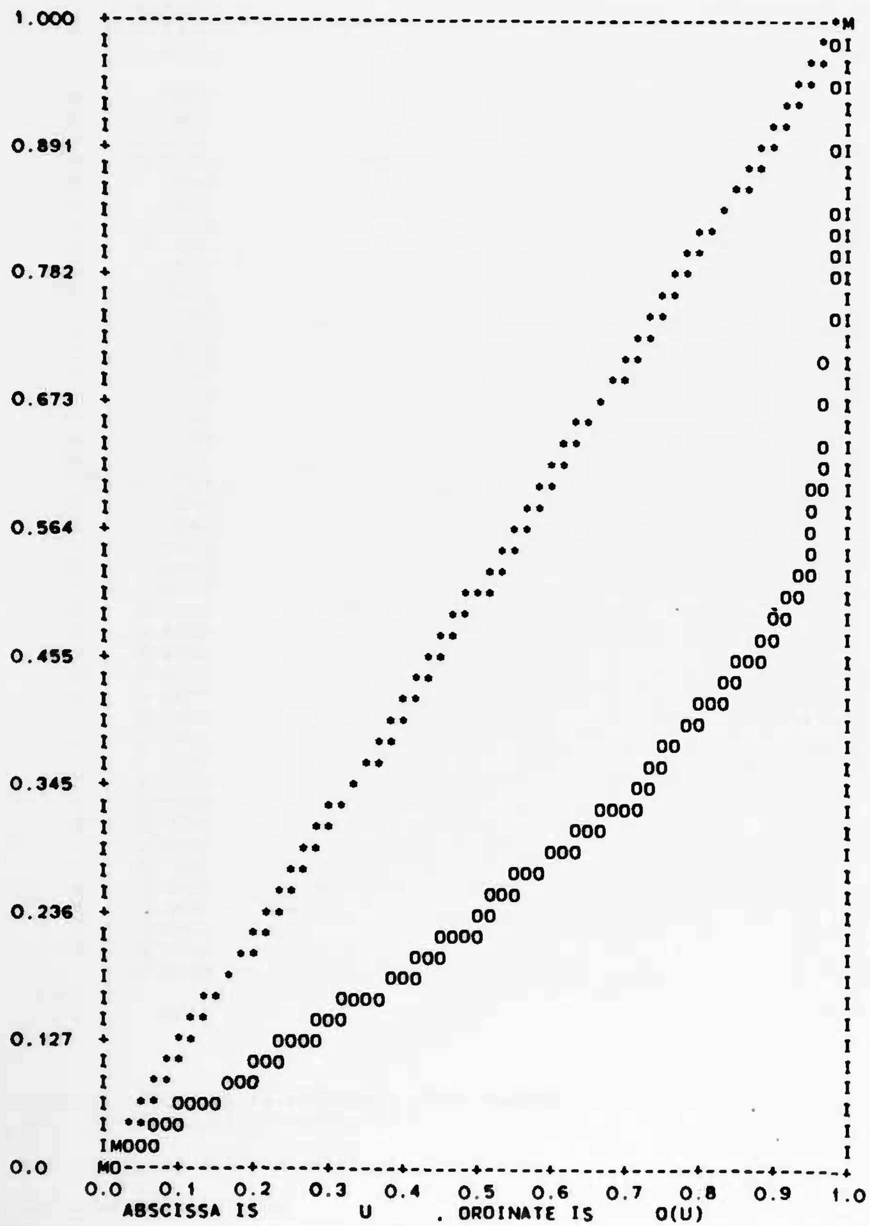


FIGURE 11

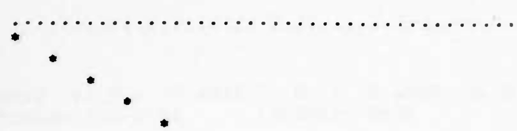
DELTA MEMORY FUNCTION

BEST ORDER AR SPECTRAL DENSITY

PLOT 1 - LAG 1 IS AT FREQUENCY 0
PLOT 2 - LAG 1 IS AT FREQUENCY KSEAS/NFREQS

I	DELT
1	0.0126
2	0.0431
3	0.0871
4	0.1415
5	0.2032
6	0.2691
7	0.3365
8	0.4038
9	0.4693
10	0.5323
11	0.5922
12	0.6487
13	0.7017
14	0.7512
15	0.7973
16	0.8401
17	0.8799
18	0.9168
19	0.9510
20	0.9826
21	1.0119
22	1.0390
23	1.0640
24	1.0871
25	1.1085
26	1.1282
27	1.1463
28	1.1630
29	1.1783
30	1.1923
31	1.2051
32	1.2168
33	1.2274
34	1.2371
35	1.2457
36	1.2535
37	1.2603
38	1.2664
39	1.2717
40	1.2762
41	1.2801
42	1.2832

I	DELT
1	0.0324
2	0.0473
3	0.0614
4	0.0748
5	0.0874



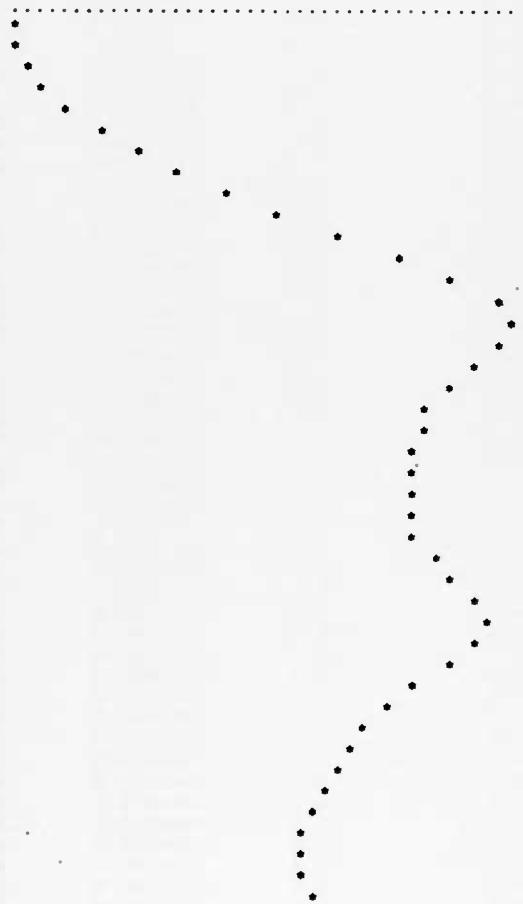
DELTA MEMORY FUNCTION

SMOOTHED PERIODOGRAM - PARZEN WINDOW

PLOT 1 - LAG 1 IS AT FREQUENCY 0
PLOT 2 - LAG 1 IS AT FREQUENCY KSEAS/NFREQS

FIGURE 12

I	DELT
1	0.0055
2	0.0213
3	0.0505
4	0.0973
5	0.1642
6	0.2511
7	0.3565
8	0.4784
9	0.6167
10	0.7726
11	0.9466
12	1.1320
13	1.3076
14	1.4353
15	1.4795
16	1.4412
17	1.3612
18	1.2846
19	1.2343
20	1.2100
21	1.1990
22	1.1889
23	1.1781
24	1.1759
25	1.1949
26	1.2410
27	1.3063
28	1.3678
29	1.3945
30	1.3670
31	1.2934
32	1.2015
33	1.1172
34	1.0518
35	1.0031
36	0.9620
37	0.9220
38	0.8846
39	0.8585
40	0.8526
41	0.8707
42	0.9074



I	DELT
1	0.0635
2	0.0851
3	0.0875
4	0.0714
5	0.0450

JOY2

LAG	RYY(V)	REY(V)	RYE(V)	REE(V)	PVH(V)
0	1.0000000	0.7326133	0.7326133	0.7326133	0.7326133
1	0.4082985	0.2064592	0.0	0.0	0.7907959
2	0.3084335	0.1225261	0.0	0.0	0.8112878
3	0.2363130	0.0475103	0.0	0.0	0.8143688
4	0.2743570	0.0961481	0.0	0.0	0.8269873
5	0.2721530	0.1046219	0.0	0.0	0.8419279
6	0.2550060	0.0854343	0.0	0.0	0.8518909
7	0.2875156	0.1448688	0.0	0.0	0.8805376
8	0.1891304	0.0624849	0.0	0.0	0.8858669
9	0.1781062	0.0541669	0.0	0.0	0.8898718
10	0.1442805	0.0158781	0.0	0.0	0.8902159
11	0.2029656	0.0737349	0.0	0.0	0.8976370
12	0.1468985	0.0206350	0.0	0.0	0.8982182
13	0.1804721	0.0577810	0.0	0.0	0.9027753
14	0.1498834	0.0220696	0.0	0.0	0.9034402
15	0.1473644	0.0215390	0.0	0.0	0.9040734
16	0.1560265	0.0296861	0.0	0.0	0.9052763
17	0.1505637	0.0043458	0.0	0.0	0.9053020
18	0.2259206	0.0951476	0.0	0.0	0.9176592
19	0.1822132	0.0607775	0.0	0.0	0.9227013
20	0.1718851	0.0420270	0.0	0.0	0.9251122
21	0.1987007	0.0762220	0.0	0.0	0.9330424
22	0.1957907	0.0851237	0.0	0.0	0.9429331
23	0.1478132	0.0375644	0.0	0.0	0.9448591
24	0.1716287	0.0677163	0.0	0.0	0.9511182
25	0.1493408	0.0537586	0.0	0.0	0.9550630
26	0.1315936	0.0387328	0.0	0.0	0.9571108
27	0.1372958	0.0449133	0.0	0.0	0.9598642
28	0.1362370	0.0478824	0.0	0.0	0.9629937
29	0.1369956	0.0565815	0.0	0.0	0.9673636
30	0.1085857	0.0297663	0.0	0.0	0.9685730
31	0.1131201	0.0373363	0.0	0.0	0.9704757
32	0.1052418	0.0316543	0.0	0.0	0.9718434
33	0.1002929	0.0275210	0.0	0.0	0.9728772
34	0.0983852	0.0275114	0.0	0.0	0.9739103
35	0.0994038	0.0282782	0.0	0.0	0.9750018
36	0.0970029	0.0281241	0.0	0.0	0.9760814
37	0.0943899	0.0285846	0.0	0.0	0.9771966
38	0.0899671	0.0227558	0.0	0.0	0.9779034
39	0.0939781	0.0277208	0.0	0.0	0.9789523
40	0.0981009	0.0361862	0.0	0.0	0.9807396
41	0.0849502	0.0229136	0.0	0.0	0.9814562
42	0.0913289	0.0312059	0.0	0.0	0.9827854
43	0.0874473	0.0310268	0.0	0.0	0.9840994
44	0.0789995	0.0237194	0.0	0.0	0.9848673
45	0.0794491	0.0252786	0.0	0.0	0.9857395
46	0.0797144	0.0280362	0.0	0.0	0.9868124
47	0.0744784	0.0253145	0.0	0.0	0.9876871
48	0.0706021	0.0226548	0.0	0.0	0.9883876
49	0.0700365	0.0238034	0.0	0.0	0.9891610
50	0.0670485	0.0230686	0.0	0.0	0.9898874
51	0.0633958	0.0207739	0.0	0.0	0.9904764
52	0.0603163	0.0189447	0.0	0.0	0.9909662
53	0.0602991	0.0206640	0.0	0.0	0.9915490
54	0.0557657	0.0172119	0.0	0.0	0.9919534
55	0.0545744	0.0172729	0.0	0.0	0.9923606
56	0.0534631	0.0171439	0.0	0.0	0.9927617
57	0.0516932	0.0162010	0.0	0.0	0.9931200
58	0.0510590	0.0171760	0.0	0.0	0.9935226

FIGURE 13

FIGURE 14

JOY2

LAG	RYY(V)	REY(V)	RYE(V)	REE(V)	PVH(V)
0	1.0000000	0.4829759	0.4829759	0.4829759	0.4829759
1	0.6747600	0.2148846	0.0	0.0	0.5785819
2	0.5941792	0.1709515	0.0	0.0	0.6390909
3	0.5314984	0.1341923	0.0	0.0	0.6763754
4	0.5120372	0.1473377	0.0	0.0	0.7213226
5	0.4531240	0.1018637	0.0	0.0	0.7428065
6	0.4374954	0.1019529	0.0	0.0	0.7643280
7	0.4114058	0.0789312	0.0	0.0	0.7772275
8	0.4142068	0.0960438	0.0	0.0	0.7963265
9	0.4060826	0.1014103	0.0	0.0	0.8176196
10	0.3925930	0.1107970	0.0	0.0	0.8430369
11	0.3698141	0.1164525	0.0	0.0	0.8711153
12	0.3145271	0.0715572	0.0	0.0	0.8817171
13	0.3090729	0.0807641	0.0	0.0	0.8952226
14	0.2859113	0.0683612	0.0	0.0	0.9048985
15	0.2759603	0.0704698	0.0	0.0	0.9151806
16	0.2572453	0.0607879	0.0	0.0	0.9228314
17	0.2483598	0.0621295	0.0	0.0	0.9308236
18	0.2345242	0.0569671	0.0	0.0	0.9375428
19	0.2263170	0.0598399	0.0	0.0	0.9449568
20	0.2116776	0.0547993	0.0	0.0	0.9511744
21	0.1994585	0.0525390	0.0	0.0	0.9568897
22	0.1859488	0.0471915	0.0	0.0	0.9615008
23	0.1746851	0.0428935	0.0	0.0	0.9653102
24	0.1670123	0.0425371	0.0	0.0	0.9690565
25	0.1570423	0.0390402	0.0	0.0	0.9722122
26	0.1494720	0.0379307	0.0	0.0	0.9751911
27	0.1408362	0.0351630	0.0	0.0	0.9777511
28	0.1339873	0.0342218	0.0	0.0	0.9801759
29	0.1262708	0.0319888	0.0	0.0	0.9822945
30	0.1196540	0.0308438	0.0	0.0	0.9842642
31	0.1123386	0.0283759	0.0	0.0	0.9859313
32	0.1062982	0.0270292	0.0	0.0	0.9874440
33	0.1001922	0.0251416	0.0	0.0	0.9887527
34	0.0949577	0.0240069	0.0	0.0	0.9899460
35	0.0898069	0.0227374	0.0	0.0	0.9910164
36	0.0848550	0.0214458	0.0	0.0	0.9919686
37	0.0802849	0.0203648	0.0	0.0	0.9928273
38	0.0758325	0.0192040	0.0	0.0	0.9935908
39	0.0717288	0.0182451	0.0	0.0	0.9942800
40	0.0676939	0.0171597	0.0	0.0	0.9948897
41	0.0639889	0.0162537	0.0	0.0	0.9954367
42	0.0603912	0.0152592	0.0	0.0	0.9959188
43	0.0571357	0.0144901	0.0	0.0	0.9963534
44	0.0539653	0.0136472	0.0	0.0	0.9967390
45	0.0510500	0.0129597	0.0	0.0	0.9970868
46	0.0482171	0.0122196	0.0	0.0	0.9973959
47	0.0455774	0.0115674	0.0	0.0	0.9976729
48	0.0430619	0.0109281	0.0	0.0	0.9979202
49	0.0406882	0.0103310	0.0	0.0	0.9981411
50	0.0384416	0.0097622	0.0	0.0	0.9983384
51	0.0363132	0.0092182	0.0	0.0	0.9985143
52	0.0343122	0.0087145	0.0	0.0	0.9986715
53	0.0324147	0.0082308	0.0	0.0	0.9988118
54	0.0306326	0.0077892	0.0	0.0	0.9989374
55	0.0289337	0.0073559	0.0	0.0	0.9990494
56	0.0273375	0.0069600	0.0	0.0	0.9991496
57	0.0258166	0.0065718	0.0	0.0	0.9992390
58	0.0243873	0.0062160	0.0	0.0	0.9993190

12. Concluding Remarks

It is important to understand the role of memory when using [for time series model identification] ARIMA (p,d,q) models introduced by Box and Jenkins (1970). Memory is related to d, but not to the orders p and q. An AR(1) process $Y(t)$ satisfying $g_1(L)\ Y(t) = \varepsilon(t)$ where $g_1(z) = 1-\rho z$ is diagnosed as long memory when the transfer function $g_1(z)$ has its root $1/\rho$ close to the unit circle in the complex z-plane. An example of a long memory population correlation function is $\rho(v) = \cos 2\pi\omega t$, which can be regarded as corresponding to an AR(2) scheme whose transfer function $g_2(z) = 1-(2 \cos 2\pi\omega)z + z^2$ has roots on the unit circle. In the ARSPIQ approach to time series model identification, roots are not explicitly evaluated because their role is subsumed by memory.

The models automatically identified by ARSPIQ have been found in practice to have the same quality as exact models for purposes of forecasting and spectral estimation. Other diagnostics of model structure (such as correlations, partial correlations, and inverse correlations) are also generated in ARSPIQ and can be used in traditional ways to guess model structure.

There are still many open problems in the theory of time series model identification, such as tests to determine which of several possible models fits best. FUN.STAT (statistical reasoning based on quantiles, entropy and information, and functional statistical inference) may be able to help statistical scientists find better solutions to problems of model identification.

REFERENCES

A. Empirical Time Series Modeling

Parzen, Emanuel (1964) "An approach to empirical time series Analysis" J. Res. Nat. Bur. Standards, Sect. D, 68D, 937-951.

Parzen, Emanuel (1969) "Multiple time series modeling" Multivariate Analysis - II, edited by P. Krishnaiah, Academic Press: New York, 389-409.

Parzen, Emanuel (1974) "Some Recent Advances in Time Series Modeling" IEEE Transactions on Automatic Control, AC-19, 723-730.

Parzen, Emanuel (1977) "Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes" Multivariate Analysis - IV, Edited by P. Krishnaiah, North Holland: Amsterdam, 283-295.

Parzen, Emanuel (1979) "Forecasting and Whitening Filter Estimation" TIMS Studies in the Management Sciences, 12, 149-165.

Parzen, Emanuel (1980) "Time Series Modeling, Spectral Analysis, and Forecasting" Time Series Analysis, ed. D. R. Brillinger and G. C. Tiao, Institute of Mathematical Statistics, 80-111.

Parzen, Emanuel (1981) "Time Series Model Identification and Prediction Variance Horizon," Proceedings of Second Tulsa Symposium on Applied Time Series Analysis. Academic Press: New York, p. 415-447.

Parzen, Emanuel (1982) "ARARMA Models for Time Series Analysis and Forecasting", Journal of Forecasting, 1, 67-82.

Parzen, Emanuel (1983a) "Time Series Model Identification by Estimating Information" Studies in Econometrics, Time Series, and Multivariate Statistics in Honor of T. W. Anderson, ed. S. Karlin, T. Amemiya, L. Goodman, Academic Press.

Parzen, Emanuel (1983b) "Time Series ARMA Model Identification by Estimating Information" Proceedings of the 15th Annual Symposium on the Interface of Computer Science and Statistics, Amsterdam: North Holland.

Parzen, Emanuel (1983c) "Time Series Model Identification, Spectral Estimation, and Functional Inference" Signal Processing in the Ocean Environment ONR Workshop Proceedings, ed. E. J. Wegman. Marcel Dekker: New York.

B. Quantile and FUN.STAT Data Modeling

Parzen, Emanuel (1979) "Nonparametric Statistical Data Modeling" Journal of the American Statistical Association, (with discussion), 74, 105-131.

Parzen, Emanuel (1982) "Data Modeling Using Quantile and Density-Quantile Functions", Some Recent Advances in Statistics. ed. J. T de Oliveira and B. Epstein, Academic Press: New York, 23-52.

Parzen, Emanuel (1983a) "Quantiles, Parametric-Select Density Estimation, and Bi-Information Parameter Estimators," Proceedings of the 14th Annual Symposium on the Interface of Computer Science and Statistics, New York: Springer Verlag, 241-245.

Parzen, Emanuel (1983b) "Entropy Interpretation of Goodness of Fit Tests," Proceedings of the 28th Army Conference on the Design of Experiments (Monterey).

Parzen, Emanuel (1983c) "FUN.STAT Quantile Approach to Two Sample Statistical Data Analysis." The Canadian Journal of Statistics.

Parzen, Emanuel (1983d) "Informative Quantile Functions and Estimation of Tail Behavior of Distributions," Manuscript.

C. Long Memory and/or Long Tailed Time Series

Freedman, L. and Lane, D. (1981) The empirical distribution of the Fourier coefficients of a sequence of independent, identically distributed long-tailed random variables Z. Wahrscheinlichkeits theorie verw. Gebiete, 55, 21-37.

Geweke, J. and Porter-Hudak, S. (1983) The Estimation and Application of Long Memory Time Series Models. Journal of Time Series Analysis.

Granger, C. W. G. and Joyeux, R. (1980) An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis, 1(1), 15-29.

Hannan, E. J. and Kanter, M. (1977) Autoregressive processes with infinite variance, J. Appl. Prob., 14, 411-415.

Hosking, J. R. M. (1981) Fractional differencing. Biometrika, 68(1), 165-176.

Janacek, G. J. (1982) Determining the degree of differencing for time series via the log spectrum. Journal of Time Series Analysis, 3, 177-183.

Mandelbrot, B. (1973) Statistical methodology for nonperiodic cycles: from the covariance to R/S analysis. Review of Economic and Social Measurement, pp. 259-290.

Mandelbrot, B. (1982) The Fractal Geometry of Nature, Freeman: San Francisco.

Rosenblatt, M. (1981) Limit theorems for Fourier transforms of functionals of Gaussian sequences, Z Wahrscheinlichkeits theorie verw. Gebiete, 55, 123-132.

D. General Empirical Time Series Analysis References

Akaike, H. (1974) A new look at the statistical model identification. IEEE Trans Autom Control, AC-19, 761-723.

Akaike, H. (1977) On entropy maximization principle. Applications of Statistics, P. R. Krishnaiah, ed. North Holland: Amsterdam, 27-41.

Box, G. E. P. and Jenkins, G. M. (1970) Time Series Analysis, Forecasting, and Control. Holden Day: San Francisco.

Hannan, E. J. (1980) The estimate of the order of an ARMA process. Ann Statist, 10, 1071-1080.

Priestley, M. B. (1981) Spectral Analysis and Time Series, Academic Press: London.

Schuster, A. (1898) On the investigation of hidden periodicities with applications to a supposed 26-day period of meteorological phenomena" Terr. Magn. 3, 13-41.

Wiener, N. (1930) Generalized harmonic analysis. Acta Math, 5, 117-258.

Yule, G. U. (1927) On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. Philos. Trans. Roy. Soc. London, Ser. A; 226, 267-298.

APPENDIX

QUANTILE MEMORY ANALYSIS OF SIMULATED AR(1)

I_∞	ρ	DATA LNSDIQ	DATA LNSGMO	MEDIAN	VAR	CORR MS	AIC ORDER	CAT ORDER	ARMA AR ORDER	PREDICTION HORIZON
.020	.2	-1.06	-1.06	.71	1.26	.004	1	1	0	2
.047	.3	- .96	- .96	.55	1.73	.007	1	1	1	2
.087	.4	-1.02	-1.02	.52	1.66	.006	1	1	1	2
.144	.5	- .94	- .94	.46	1.91	.007	2	2	1	2
.223	.6	- .92	- .92	.50	1.87	.007	1	1	1	2
.337	.7	- .96	- .96	.30	3.10	.013	2	2	1	3
.511	.8	-1.06	-1.06	.18	14.9	.065	1	1	1	10
.830	.9	-1.16	-1.16	.08	22.3	.11	1	1	1	22
1.164	.95	-1.07	-1.08	.06	24.7	.11	10	10	1	73
2	.991	- .92	- .95	.02	46.8	.22	1	1	1	82
1.75	.985	-1.12	-1.13	.014	48.7	.23	1	1	1	67
1.5	.975	- .77	- .79	.05	24.0	.09	2	2	1	22
1.25	.958	-1.14	-1.15	.02	32.21	.15	1	1	1	27
1.0	.93	- .98	- .98	.04	12:09	.05	1	1	1	11
.75	.88	-1.17	-1.17	.05	6.25	.03	2	2	1	9
.50	.795	-1.16	-1.16	.27	4.73	.02	2	2	1	3
.25	.627	-1.00	-1.01	.33	3.36	.01	1	1	1	3

ARSPIQ

The ARSPIQ Fortran Computer Program for Time Series Model Identification by estimating information and memory is used at Texas A&M in a batch mode. It generates the following output for examination by the time series analyst.

1. Quantile data analysis of original data: $\tilde{IQ}(u)$
   Goodness of fit of normal distribution: D(u).
   LNSQID, LNSGMO
   Generates time series Y(t) with median subtracted

2. Quantile data analysis of normalized periodogram: $\tilde{IQ}(u)$
   Goodness of fit of exponential distribution: $\tilde{D}(u)$
   Median periodogram, variance periodogram
   Delta estimates at zero and seasonal frequencies (based on periodogram, usually no limit evident).

3. Quantile data analysis of correlations: $\tilde{IQ}(u)$
   Goodness of fit of normal distribution: $\tilde{D}(u)$
   Correlation mean square

4. Quantile data analysis of partial correlations: $\tilde{IQ}(u)$
   Goodness of fit of normal distribution: $\tilde{D}(u)$
   Partial correlation inter-quartile range, number of outliers

5. AR Description of time series: AIC, CAT orders
   AR coefficients for best order $\hat{m}$ and 2nd best order
   AR spectral density and spectral distribution plots

6. AR spectral density delta estimators at zero and seasonal frequencies
   Parzen window spectral density delta estimators

7. MA(∞) estimation
   AR coefficients for order $4\hat{m}$, computing partial correlations by non-stationary AR (Burg) method, or optionally by stationary AR(Yule-Walker) method
   Inverse correlations
   Infinite MA coefficients, prediction variance horizon

8. ARMA model identification by select regression
   ARMA spectral density and spectral distribution plots.

9. Cepstral pseudo-correlation estimation.

10. Spectral local quantile estimation.

DAT

ILM

process. Ann Statist,

Priestley, M. B. (1981) Spectral Analysis and Time Series, Academic Press: London.

Schuster, A. (1898) On the investigation of hidden periodicitie[illegible] with applications to a supposed 26-day period of meteorological phenomena" Terr. Magn. 3, 13-41.

I_∞	ρ
.020	.2
.047	.3
.087	.4
.144	.5
.223	.6
.337	.7
.511	.8
.830	.9
1.164	.95
2	.991
1.75	.985
1.5	.975
1.25	.958
1.0	.93
.75	.88
.50	.795
.25	.627

9. Cepstral pseudo-correlation estimation.
10. Spectral local quantile estimation.

DTIC