

AD-A132 240

TIME SERIES LONG MEMORY IDENTIFICATION AND QUANTILE
SPECTRAL ANALYSIS(U) TEXAS A AND M UNIV COLLEGE STATION
DEPT OF STATISTICS E PARZEN AUG 83 TR-A-23

1/1

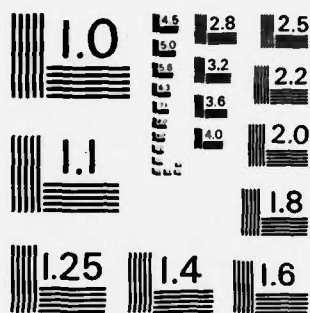
UNCLASSIFIED

ARO-20140.6-MA DAAG29-83-K-0051

F/G 12/1

NL





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

12

TEXAS A&M UNIVERSITY

COLLEGE STATION, TEXAS 77843-3143

STATISTICS
Phone 409 - 845-3141



TIME SERIES LONG MEMORY IDENTIFICATION AND QUANTILE SPECTRAL ANALYSIS

by Emanuel Parzen
Department of Statistics
Texas A&M University

Technical Report No. A-23

August 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

Approved for public release; distribution unlimited.

83 09 06 019

ADA 132240

DTIC FILE COPY

DTIC
SELECTED
SEP 08 1983
E

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER A-23	2. GOVT ACCESSION NO. AD A132 240	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Time Series Long Memory Identification and Quantile Spectral Analysis		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 Institute 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 August 1983
		13. NUMBER OF PAGES
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, quantile, spectral analysis, model identification, long memory time series, spectral estimation, subset regression, ARMA identification.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) An approach to spectral estimation is described which involves the simultaneous use of frequency, time and quantile domain algorithms, and is called <u>quantile spectral analysis</u> . It is based on the premise that while the spectrum is a non-parametric concept, its estimation cannot be a non-parametric procedure to be conducted independently of model identification. We discuss: the goals of spectral analysis, quantile data analysis, identification of memory (no, short, long), index of regular variation of a spectral		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 68 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

density, autoregressive spectral estimation, and ARMA model identification by estimating $MA(\infty)$ and subset regression. An illustrative example is given of quantile spectral analysis.

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	



S/N 0102- LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TIME SERIES LONG MEMORY IDENTIFICATION
AND QUANTILE SPECTRAL ANALYSIS

by

Emanuel Parzen
Department of Statistics
Texas A&M University
College Station, TX 77843
409-845-3188

Abstract

An approach to spectral estimation is described which involves the simultaneous use of frequency, time, and quantile domain algorithms, and is called quantile spectral analysis. It is based on the premise that while the spectrum is a non-parametric concept, its estimation cannot be a non-parametric procedure to be conducted independently of model identification. We discuss: the goals of spectral analysis, quantile data analysis, identification of memory (no, short, long), index of regular variation of a spectral density, autoregressive spectral estimation, and ARMA model identification by estimating $MA(\phi)$ and subset regression. An illustrative example is given of quantile spectral analysis.

infinity

Research supported in part by the Army Research Office and Office of Naval Research under contracts DAAG29-83-K-0051, ONR N00014-83-WRM3008.

-a-

1. Introduction to a theory of spectral synthesis

Statistical Spectral Analysis appears to be a subject of considerable controversy as to how to do it and whether to do it. In many fields of engineering and physical sciences, its importance for applications is well recognized. In other fields (notably economics) its value is still debated. One reason for this may be the difficulty of analysis of time series with trends or very slowly decaying correlations or very low frequency cycles or spectral densities with very large dynamic range. A single name for such time series is "long memory" time series.

This paper describes an approach to time series analysis which attempts to use simultaneously diverse domains of analysis, and thus to meet the needs of all the possible fields of application of time series analysis. It also aims to integrate spectral and correlation methods with methods for long memory and/or long tailed time series.

The correlation function and spectrum are basic non-parametric (or functional) parameters used to model and data analyze time series. Estimation of the correlation function and of the spectrum represent two of the basic tools used for descriptive data summaries of observations and to guess parametric probability models to fit to observations. The spectrum is important also as a major concept in terms of which to analyze the effect of passing random processes (representing either signal or noise) through linear (and, to some extent, non-linear) systems.

Correlation and spectrum are examples of non-parametric signatures of parametric models. We believe that such signatures provide key (and two-key) methods for achieving the goals of time series analysis (and statistical data analysis). The goals are to find: "Theories to fit (attest) the (statistical) facts" and "statistical facts to fit (test) theories." By fitting theories to facts one means either statistical models (to describe the statistical behavior of the data) or scientific models (to explain the statistical models fitted by the data). By statistical facts to test theories one means the estimation of characteristics of non-parametric statistical models (significant time lags, significant frequencies, and memory); such parameters (estimated non-parametrically) represent descriptions of a real process which an acceptable (or parametric model) must explain. The goals of time series analysis can be stated simply: seek models which fit curves (or fit samples), where fit is measured by the degree of scientific insight provided into underlying physical mechanisms.

The approach to spectral estimation described in this paper involves the simultaneous use of diverse algorithms for time series analysis (it could be called spectral synthesis). Our approach is based on a premise that might appear paradoxical: while the spectrum is a non-parametric concept, its estimation cannot be a non-parametric procedure to be conducted independently of model identification.

To form a spectral estimator one must identify the memory type of the time series, which we classify into one of 3 types:

- a. No memory or white noise,
- b. Short memory or stationary with finite spectral dynamic range,
- c. Long memory.

A short memory time series is modeled parametrically by the invertible filters which transforms it to white noise whose type (AR, MA, or ARMA) one must identify.

A long memory time series is modeled parametrically by an operator which transforms it to a short memory time series; such operators are non-invertible filters or representations as the sum of a long-memory signal and a short memory noise.

The goal of the time series analyst is often defined to be either a time domain model or a spectral analysis. Our approach maintains that the two domains must be employed simultaneously because the choice of final answer must be based on having a satisfactory interpretation in both domains. Additional domains (involving memory, information, and quantiles) are utilized in our approach to time series model identification, especially new diagnostic measures (or model signatures), based on "quantile data analysis" of spectral density and correlation functions. These new model signatures represent an application to time series analysis of new time-series theoretic methods of

statistical data analysis of probability distributions which we call Quantile Data Analysis and Functional Statistical Inference (abbreviated FUN.STAT).

The FUN.STAT approach to statistical data analysis is based on isomorphisms between properties of spectral density functions and density-quantile functions. One of the rewards of this isomorphism is an important diagnostic of time series memory called the index δ of regular variation of a spectral density at frequency ω .

2. How to define the spectrum

As the goal of the theory of spectral analysis, we propose that we adopt the goal stated by Wiener (1930) in his celebrated pioneering paper which introduced generalized harmonic analysis. Wiener defined the goal of spectral analysis to be: to improve, and make rigorous, Schuster's concept of the periodogram of a sample. We consider only discrete parameter time series $Y(t)$, $t=0, \pm 1, \dots$. A sample is the finite (but increasing) number of observations $Y(t)$, $t=1, 2, \dots, T$.

To detect the "hidden periodicity" ω_0 in the model

$$(1) \quad Y(t) = A \cos \omega_0 t + B \sin \omega_0 t + N(t),$$

where $N(t)$ is white noise [a sequence of independent identically distributed random variables with finite second (and possibly higher) moments], Schuster (1898) proposed calculating the function $S_T(\omega)$, $-0.5 \leq \omega \leq 0.5$ [we take $0 \leq \omega \leq 1$], defined by

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

which we call the sample unnormalized spectral density or periodogram or sample power spectrum.

When the time series obeys the model (1), one can show that $S_T(\omega_0)$ tends to ∞ as T tends to ∞ . Therefore one might interpret

local maxima of $S_T(\omega)$ as indicating "significant frequencies" representing "hidden periodicities." However the graph of $S_T(\omega)$ is often a very wiggly function, and one obtains many bumps in the spectrum representing "spurious periodicities."

A traditional approach in statistical communication engineering textbooks to defining the spectrum $S(\omega)$ of a time series $Y(t)$, $t=0, \pm 1, \dots$ has been

$$(3) \quad S(\omega) = \lim_{T \rightarrow \infty} S_T(\omega) \quad .$$

Unfortunately for those who would like the world to be simple [but fortunately for those who enjoy the deeper beauty of a more complicated reality and the accompanying theory] the limit of $S_T(\omega)$ does not exist in any usual mode of convergence. The story of spectral analysis starts with the study of the limiting behavior of $S_T(\omega)$, $-0.5 \leq \omega \leq 0.5$, especially how to use it to define statistically significant signatures of time series samples. To solve the problem of interpreting $S_T(\omega)$, Wiener (1930) proposed a "radical recasting" based on the Fourier representations

$$(4) \quad \begin{aligned} S_T(\omega) &= \sum_{v=-T}^T \exp(-2\pi i v \omega) R_T(v) \quad , \\ R_T(v) &= \int_{-0.5}^{0.5} \exp(2\pi i v \omega) S_T(\omega) d\omega \end{aligned}$$

where $R_T(v)$ is the sample covariance function

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

The sample correlation function is defined by

$$(6) \quad \rho_T(v) = R_T(v) \div R_T(0)$$

Wiener's theory of generalized harmonic analysis is based on assuming the existence of the limit

$$(7) \quad \rho(v) = \lim_{T \rightarrow \infty} \rho_T(v) .$$

One calls $\rho(v)$ the (asymptotic) correlation function.

One can show that $\rho(v)$ is a function of non-negative type: for any set of complex numbers $c_1, \dots, c_n$ and integers $v_1, \dots, v_n$

$$(8) \quad \sum_{i,j=1}^n c_i c_j^* \rho(v_i - v_j) \geq 0$$

where c_j^* denotes the complex conjugate of c_j . Consequently, there exists a bounded non-decreasing function of a real variable ω , denoted $F(\omega)$, and called the spectral distribution function,

such that

$$(9) \quad \rho(v) = \int_{-0.5}^{0.5} e^{2\pi i v \omega} dF(\omega), \quad v=0, \pm 1, \dots$$

The spectral density function $f(\omega)$ is defined as the derivative of the absolutely continuous part of $F(\omega)$. If $F(\omega)$ is itself absolutely continuous, then

$$\rho(v) = \int_0^1 e^{2\pi i v \omega} f(\omega) d\omega.$$

$$(10) \quad f(\omega) = \sum_{v=-\infty}^{\infty} e^{-2\pi i v \omega} \rho(v).$$

A sufficient condition for the spectral density $f(\omega)$ to exist is that the correlation function $\rho(v)$ tend to 0 (as $v \rightarrow \infty$) sufficiently fast that

$$(11) \quad \sum_{v=-\infty}^{\infty} |\rho(v)| \quad \text{or} \quad \sum_{v=-\infty}^{\infty} |\rho(v)|^2 < \infty.$$

A probability model under which the asymptotic correlation function $\rho(v)$ may be proved to exist is the following:
 $Y(t)$, $t=0, \pm 1, \dots$ is a zero mean covariance stationary time series with covariance function

$$(12) \quad R(v) = E[Y(t+v) Y(t)]$$

and correlation function $\rho(v) = R(v) \div R(0)$. When the time series is Gaussian, a necessary and sufficient condition for $\rho_T(v)$ to converge (almost surely) to $\rho(v)$ is that

$$(13) \quad \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{v=1}^T \rho^2(v) = 0.$$

This is another example of a condition on $\rho(v)$ which reflects the rate of decay to zero of $\rho(v)$.

A necessary and sufficient condition for a zero mean Gaussian covariance stationary time series to be white noise is that $\rho(v) = 0$ for $v \neq 0$. It is natural to conclude that one can distinguish three types of time series:

$$\text{no memory} \quad \frac{1}{T} \sum_{v=1}^T \rho^2(v) = 0 \quad \text{for all } T$$

$$\text{short memory} \quad \frac{1}{T} \sum_{v=1}^T \rho^2(v) \rightarrow 0 \quad \text{as } T \rightarrow \infty$$

$$\text{long memory} \quad \frac{1}{T} \sum_{v=1}^T \rho^2(v) \neq 0 \quad .$$

One can argue that: (1) an optimal non-parametric estimator $\hat{\rho}(v)$ of $\rho(v)$ is given by $\hat{\rho}(v) = \rho_T(v)$; and (2) one can base a test of the hypothesis H_0 of white noise on a suitable number of sample correlations $\hat{\rho}(v)$, $v=1,2,\dots,n$. Under H_0 , the asymptotic distribution of $\rho_T(v)$ is independent Gaussian random variables with zero means and variance T^{-1} . To test H_0

one can test whether $\hat{\rho}(v)$, $v=1, \dots, n$, satisfy the hypothesis that they are a random sample from a normal distribution with mean 0 and variance $1/T$. This is the kind of hypothesis for which FUN.STAT provides tools.

A quick and dirty diagnostic of time series memory type is provided by the value of the correlations mean square

$$\text{CORRMS} = \frac{1}{n} \sum_{v=1}^n \hat{\rho}^2(v) \quad ..$$

Computation of CORRMS for samples of size 200 indicate that very approximately $0.004 \leq \text{CORRMS} \leq 0.1$ indicates short memory (then $\text{CORRMS} \leq 0.004$, no memory; $\text{CORRMS} \geq .1$, long memory).

The sample spectral density

$$\tilde{f}(\omega) = \sum_{v=-T}^T \exp(-2\pi i v \omega) \hat{\rho}(v)$$

is computed by first computing the sample Fourier transform

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

at an equi-spaced grid of frequencies in $0 \leq \omega \leq 1$ of the form $\omega = k/S$, $k=0, 1, \dots, S-1$. 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 $\hat{\rho}(v)$.

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:

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

When a time series is no memory (or white noise), and has finite second moments, it is a basic theorem of time series analysis that asymptotically the random variables $f(k/S)$, $k=1, \dots, [S/2]$, are identically distributed as an exponential distribution with mean 1. Therefore tests for white noise can be obtained by quantile data analysis based tests for exponentiality of the sample spectral density $f(\omega)$ at suitable frequencies (if the informative quantile function of the original data does not indicate a long tailed distribution).

Powerful discriminators of memory type are SPECMED (the median) and SPECVAR (the variance) of the data batch of values of the sample spectral density. For no memory these are the median and variance of an exponential distribution with mean 1:

$$\text{SPECMED} = \log 2 = .69, \quad \text{SPECVAR} = 1.$$

Computation of these signatures for samples of size 200 indicate that short memory corresponds very approximately to

$$.1 \leq \text{SPECMED} \leq .6, \quad 1.5 \leq \text{SPECVAR} \leq 20 .$$

Deeper diagnostics of time series memory are provided by quantile data analysis of $f(\omega)$, especially plots of informative quantile function $\tilde{I}Q(u)$ and comparison distribution functions $\tilde{D}(u)$.

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 dynamic range of the spectral density function and sample spectral density, defined

$$\text{SPECRNG} = \left\{ \max_{\omega} \log f(\omega) - \min_{\omega} \log f(\omega) \right\} .$$

Dynamic range classification of memory of a time series:

no memory $\equiv$ dynamic range = 0 ,

short memory $\equiv$ $0 < \text{dynamic range} < \infty$,

long memory $\equiv$ dynamic range = ∞ .

When the dynamic range is finite, we can assume that the spectral density $f(\omega)$ is bounded above and below: for some constants c_1 and c_2 , $0 < c_1 \leq f(\omega) \leq c_2 < \infty$ for all ω .

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 can be defined in terms of the index of regular variation of the spectral density at a specified frequency, usually zero frequency. As $\omega \rightarrow 0$, the spectral density $f(\omega)$ is assumed to be a regularly varying function, with the representation

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

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

$$\text{No and short memory} \quad \equiv \quad \delta = 0$$

$$\text{Long memory} \quad \equiv \quad \delta \neq 0 \quad .$$

Long memory time series models have spectral density $f(\omega)$ satisfying the regular variation representation. The index $\delta < 0$ ($\delta > 0$) corresponds to a zero (infinite) value for $f(\omega)$ at $\omega = 0$.

Determining the degree of differencing: 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 $\delta = 2d$. Estimators for δ can provide techniques for estimating d .

Self-similarity: When a spectral density $f(\omega)$ is regularly varying at $\omega = 0$ it enjoys a property called approximate self-similarity: $f(y\omega) = y^{-\delta} f(\omega)$ in the sense that, for any $y > 0$, $f(y\omega) \div y^{-\delta} f(\omega) \rightarrow 1$ as $\omega \rightarrow 0$.

3. Quantile Data Analysis

The probability distribution of a random variable X has been traditionally described by its distribution function $F(x) = \Pr[X \leq x]$ and its probability density function $f(x) = F'(x)$. Given a sample $X_1, \dots, X_n$ of X , the models that we seek to fit to the data are usually parametric models of the form

$$F(x) = F_0\left(\frac{x-\mu}{\sigma}\right)$$

for parameters μ and σ (representing location and scale respectively) to be estimated, and $F_0(x)$ a known distribution function. The most important cases of $F_0(x)$ are:

$$\text{normal} \quad F_0(x) = \Phi(x) = \int_{-\infty}^x \phi(y) dy \quad ,$$

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

$$\text{exponential} \quad F_0(x) = 1 - e^{-x} \quad , \quad x > 0 \quad .$$

The quantile function $Q(u)$, $0 \leq u \leq 1$, defined by

$$Q(u) = F^{-1}(u) = \inf \{x: F(x) \geq u\} \quad ,$$

can be regarded as inverse of the distribution function. When $F(x)$ is continuous, $FQ(u) = u$. Quantile data analysis estimates

non-parametrically $Q(u)$; the quantile-density function $q(u) = Q'(u)$; and the density-quantile function

$$fQ(u) = f(F^{-1}(u)) = \{q(u)\}^{-1} .$$

The tail behavior of a distribution can be described by indices α_0 and α_1 in the following representations of the density-quantile function as a regularly varying function:

$$fQ(u) = u^{\alpha_0} L_0(u), \quad fQ(u) = (1-u)^{\alpha_1} L_1(u)$$

where $L_j(u)$ is a slowly varying function as $u \rightarrow j$.

The definition of a function $L(u)$ slowly varying at $u=0$ is: for every y , in $0 < y < 1$,

$$\log L(yu) - \log L(u) \rightarrow 0 \quad \text{as} \quad u \rightarrow 0 .$$

We assume in addition that

$$\lim_{u \rightarrow 0} \int_0^1 \{\log L(yu) - \log u\} dy = 0 .$$

Then one can compute α_0 by

$$-\alpha_0 = \lim_{u \rightarrow 0} \int_0^1 \{\log fQ(yu) - \log fQ(u)\} dy .$$

We digress to note that for a spectral density function $f(\omega)$, $0 \leq \omega \leq 1$ one assumes a representation (at $\omega=0$)

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

We call δ the index of regular variation at frequency $\omega=0$. It is computed by

$$\delta = \lim_{\omega \rightarrow 0} \int_0^1 \{ \log f(y\omega) - \log f(\omega) \} dy \quad .$$

Estimation of the density-quantile function $fQ(u)$ can be treated by similar procedures as are used to estimate spectral densities $f(\omega)$. However much insight into the distributions that might fit a sample can be obtained by non-parametrically estimating the quantile function and the amazingly useful informative quantile function

$$IQ(u) = \frac{Q(u) - \mu_1}{\sigma_1}$$

where μ_1 and σ_1 are universal estimators of location and scale respectively. We propose $\mu_1 = Q(0.5)$, $\sigma_1 = Q'(0.5) = q(0.5)$.

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). A plot of $IQ(u)$ is accompanied by its values at $u = 0.01, 0.05, 0.10, 0.25, 0.75, 0.90, 0.95, 0.99$.

Natural quick and dirty estimators of σ_1 are

$$\sigma_p = \{Q(0.5 + p) - Q(0.5 - p)\} \div 2p \quad ;$$

where $0 \leq p \leq 0.5$; our preferred estimator is $\sigma_{0.25}$ which equals twice the interquartile range:

$$\sigma_{.25} = 2 \{Q(0.75) - Q(0.25)\} \quad .$$

To estimate these non-parametric parameters from a sample of size n , we form a sample quantile function $\tilde{Q}(u)$ by linear interpolation of the values

$$\tilde{Q}\left(\frac{j}{n+1}\right) = X_{jn} \quad , \quad j = 1, \dots, n$$

where $X_{1n} \leq \dots \leq X_{nn}$ are the order statistics of the sample.

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

One can test (before parameter estimation) the goodness of fit of a sample to $F(x) = F_0(\frac{x-\mu}{\sigma})$ for some μ and σ by introducing the weighted spacings

$$\tilde{d}(u) = \frac{1}{\tilde{\sigma}_0} f_0 Q_0(u) \tilde{q}(u)$$

where: $f_0 Q_0(u) = f_0(F_0^{-1}(u))$ 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

$$\tilde{\sigma}_0 = \int_0^1 f_0 Q_0(u) \tilde{q}(u) du$$

The test function is

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

which one compares with the uniform distribution $D(u) = u$, $0 \leq u \leq 1$. We call $\tilde{D}(u)$ the sample comparison distribution function, or the cumulative weighted spacings function [Parzen (1979)].

The data batch $\tilde{f}(\frac{k}{S})$, $k=0, 1, \dots, S/2$, is tested for exponentiality by forming its informative quantile function $\tilde{IQ}(u)$ and its cumulative weighted spacings function $\tilde{D}(u)$, with $f_0 Q_0(u) = 1-u$. How one interprets the quantile data analysis of the sample spectral density is best illustrated by examples.

4. Autoregressive spectral estimation

The concept of an autoregressive representation of a time series can be defined from several points of view. When one's goal is spectral estimation based on improving Schuster's definition of the spectrum, it seems natural to adopt the viewpoint of a deconvolution filter representation:

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

One seeks to choose the order p and coefficients $a_p(j)$ so that the time series $\epsilon(t)$ is parsimoniously white noise (just barely passes tests for white noise). From the deconvolution filter representation of $Y(t)$ one obtains approximately the following formula for sample spectral density functions:

$$\tilde{f}_Y(\omega) \left| \sum_{j=0}^p a_p(j) \exp(2\pi i \omega j) \right|^2 = \hat{\sigma}_p^2 \tilde{f}_\epsilon(\omega)$$

where

$$\hat{\sigma}_p^2 = \frac{\sum_{t=1}^T \epsilon^2(t)}{\sum_{t=1}^T Y^2(t)}$$

It is important to note that a near zero value of $\hat{\sigma}_p^2$ is regarded as evidence that memory type is long.

When $\tilde{f}_\epsilon(\omega)$ behaves as the sample spectral density of white noise, we can estimate it by the constant 1. As potential

estimates of the spectral density $f_Y(\omega)$ of the time series $Y(t)$ we consider the sequence of autoregressive spectral estimators

$$\hat{f}_p(\omega) = \hat{\sigma}_p^2 \left| \sum_{j=0}^p a_p(j) \exp(2\pi i \omega j) \right|^{-2}$$

for $p=0,1,2,\dots$ with coefficients $a_p(j)$ computed by suitable algorithms. The sequence $\hat{f}_p(\omega)$, $p=0,1,2,\dots$, should be regarded as a sequence of functions which smooth the sample spectral density. They become increasingly wiggly as p tends to T , and eventually coincide with $\tilde{f}_Y(\omega)$.

To estimate the autoregressive coefficients $a_p(j)$, $p=1,2,\dots$ and $j=1,2,\dots,p$, two main methods are available of which important representatives are algorithms called

1. Yule-Walker (stationary)
2. Burg (non-stationary)

The optimality properties of the methods depends on the memory type of the time series. Theoretical and empirical evidence indicate that Yule-Walker and Burg estimators agree for short memory time series (which can be shown to be always representable as an invertible infinite autoregression). For long memory time series which one assumes to possess an autoregressive representation, Yule-Walker and Burg estimators usually differ significantly; further Burg (and least squares) estimators are consistent estimators while Yule-Walker are not.

We propose the following practical consequences of these facts: given a time series sample, compute autoregressive

coefficients by both Yule-Walker and Burg estimators (for orders p to be determined from the data as described below). Check whether the two ways of estimation yield similar results. A yes answer is evidence that the memory type is short, or if memory is long that an autoregressive model may fit. If memory is short, identify an ARMA scheme by the procedures of the next section. A no answer is evidence that the memory type is long; using the diverse signatures introduced, one should determine operations which just barely transform the long memory time series to a short memory one (especially a model as a sum of long memory signal plus short memory noise).

Solutions to the important problem of determining the order of approximating autoregressive schemes can be approached using order determining criterion functions, especially AIC (due to Akaike) and CAT (due to Parzen), which are formed from the sequence $\hat{\sigma}_p^2$, $p=1,2,\dots$. We recommend that one determine two orders (called best and second best) rather than a single order. The best (second best) order is that at which the criterion function achieves its minimum (second lowest relative minimum). The maximum of these orders (denoted $\hat{p}$) is used as the order for least squares (or Burg) estimation of autoregressive coefficients.

An order determining criterion should be used only to suggest autoregressive orders p for which $\hat{f}_p(\omega)$ is a satisfactory spectral estimator. The ultimate criterion of goodness of fit of an autoregressive spectral estimator is that

$$\tilde{d}_p(\omega) = \tilde{f}_Y(\omega) \div \tilde{f}_p(\omega)$$

just barely satisfies the hypothesis that it is the sample spectral density of white noise. The notation $\tilde{d}(\omega)$ is chosen to convey that this function has properties similar to that of $\tilde{d}(u)$ introduced in the preceding section for testing goodness of fit of probability distributions.

While autoregressive spectral estimation can be performed for long memory time series obeying an autoregressive scheme with roots on or very near the unit circle, it seems to me that the process of model identification of observed real time series has greatest scientific insight if it is carried out in two stages in which one first finds an autoregressive operator (with a simple interpretation) which just barely transforms the time series to short memory, which in turn is modeled by an ARMA whitening filter.

5. ARMA Model Identification by Estimating $MA(\infty)$ and Subset Regression

A variety of approaches have been proposed by statisticians for identifying the orders p and q and estimating the coefficients of an ARMA (p,q) model for an observed time series. An approach that we use (because of its very low computational cost) for an initial identification is based on first estimating the coefficients of the $MA(\infty)$ representation:

$$Y(t) = \epsilon(t) + b_{\infty}(1) \epsilon(t-1) + \dots$$

with residual variance

$$\sigma_{\infty}^2 = E[\epsilon^2(t)] \div E[Y^2(t)]$$

One approach to estimating $MA(\infty)$ is to estimate the $AR(\infty)$ representation by an approximating $AR(p)$ representation (we usually use the Burg algorithm to compute its coefficients). Then one solves recursively for $b_{\infty}(j)$ by

$$a_p(0) b_{\infty}(k) + a_p(1) b_{\infty}(k-1) + \dots + a_p(k) b_{\infty}(0) = 0, \quad k=1,2,\dots$$

A second approach to estimating $MA(\infty)$ is to compute the cepstral pseudo-correlations

$$\psi(v) = \int_0^1 \exp(2\pi i \omega v) \log f(\omega) d\omega, \quad v=0, \pm 1, \dots$$

One computes $\hat{\psi}(v)$ by replacing $f(\omega)$ by a windowed spectral density estimator

$$\hat{f}(\omega) = \sum_{v=-T}^T \exp(-2\pi i v \omega) k\left(\frac{v}{M}\right) \hat{\rho}(v)$$

for a suitable kernel $k(x)$ and truncation point M (satisfying $T/2 \leq M \leq T$).

From $\psi(v)$ one can compute $b_{\infty}(n)$ by the recursive formula

$$(n+1) b_{\infty}(n+1) = \sum_{k=0}^n (k+1) \psi(k+1) b_{\infty}(n-k)$$

The spectral formula for residual variance σ_{∞}^2 ,

$$\log \sigma_{\infty}^2 = \int_0^1 \log f(\omega) d\omega = \psi(0)$$

yields an estimator $\hat{\sigma}_{\infty}^2$ when one replaces $f(\omega)$ by $\hat{f}(\omega)$.

In the population a key relation is

$$\sigma_{\infty}^2 \{1 + b_{\infty}(1) + b_{\infty}(2) + \dots\} = 1$$

An alternative estimator of σ_{∞}^2 is therefore

$$\bar{\sigma}_{\infty}^2 = \{1 + b_{\infty}^2(1) + b_{\infty}^2(2) + \dots\}^{-1}$$

A useful signature of the memory and ARMA types of a time series

is the prediction variance horizon function

$$PVH(h) = \bar{\sigma}_{\infty}^2 \{1 + b_{\infty}^2(1) + \dots + b_{\infty}^2(h-1)\}, \quad h=1,2,\dots$$

It can be interpreted as representing the mean square error of prediction h steps ahead. The horizon of a time series is defined to be smallest value of h for which $PVH(h)$ is greater than a suitable value (such as 0.95).

From the $MA(\infty)$ representation one forms an estimator

$$\bar{p}(v) = \bar{\sigma}_{\infty}^2 \{b_{\infty}(v) + b_{\infty}(1) b_{\infty}(v+1) + \dots\}$$

of $\rho(v)$, and an estimator $\bar{\sigma}_{\infty}^2 b_{\infty}(k)$ of the covariance between $Y(t)$ and $\varepsilon(t-k)$; we assume $Y(t)$ has been normalized to have variance 1.

Next one forms the joint covariance matrix of $Y(t)$, $Y(t-1), \dots, Y(t-m)$, $\varepsilon(t-1), \dots, \varepsilon(t-m)$ for a suitable lag m . Finally, a subset regression routine is used to determine an ARMA model

$$\begin{aligned} & a_p(0) Y(t) + a_p(1) Y(t-1) + \dots + a_p(p) Y(t-p) \\ & = b(0) \varepsilon(t) + b_q(1) \varepsilon(t-1) + \dots + b_q(q) \varepsilon(t-q) \end{aligned}$$

with as many zero coefficients as possible [note: $a_p(0) = b_q(0) = 1$]. These models, called subset regression ARMA models,

yield ARMA spectral estimates

$$\hat{f}_{p,q}(\omega) = \sigma_{p,q}^2 \frac{|h_q(e^{2\pi i\omega})|^2}{|g_p(e^{2\pi i\omega})|^2}$$

where

$$g_p(z) = a_p(0) + a_p(1)z + \dots + a_p(p) z^p,$$

$$h_q(z) = b_q(0) + b_q(1)z + \dots + b_q(q) z^q.$$

For a monthly economic time series, with short memory, one often finds $p=2$, $q=12$, with $b_{12}(12)$ the only non-zero moving average coefficient. The transfer function $g_2(z)$ models the low frequency component, and $h_{12}(z)$ models the seasonal component. We use the notation ARMA(1,2;12) for this model. We use AR(1,12,13) for a subset ARMA model with $p=13$, $q=0$, and $a_{13}(1)$, $a_{13}(12)$, $a_{13}(13)$ the only non-zero coefficients.

6. An example of quantile spectral analysis

To illustrate the quantile approach to spectral estimation, let us consider New York City monthly average temperatures 1946-1959 (such a series might be collected jointly with New York City monthly births 1947-1960 to investigate if there is a relationship between atmospheric temperature and birth rate). One suspects a seasonal period of 12 (equal to $\omega = .08333$).

Original data signatures: Mean 54.6, median 55.1

Standard deviation of the informative quantile function is .2648 with log -1.33; this diagnostic measure is -1 for Gaussian time series. The values $\tilde{I}Q(0.01) = -.48$ and $\tilde{I}Q(0.99) = .42$ provide decisive evidence that the distribution is not Gaussian, but is short tail [which in the case of time series represents a harbinger of a sine wave plus noise model].

Sample Spectral Density $\tilde{f}$: Median .06, variance 50 are strong evidence that the time series is long memory. Quantile density $\tilde{q}(u)$ of $\tilde{f}(\omega)$ has maximum value 30892; extreme values of quantile $\tilde{Q}(u)$ of $\tilde{f}$ are 25 and 79 [such large values indicate the presence of a very narrow band signal]. The graph of $\tilde{D}(u)$ confirms this conclusion.

Correlations. Mean square .26 is strong evidence that time series is long memory with sine wave components.

AR order determination. As usual, the same best and second best AR orders are reported by AIC and CAT. The orders are 9 and 7, with $\hat{\sigma}_p^2$ equal to .097 and .099 respectively.

Delta (index of regular variation) at $\omega=0$, 0.08333:

Autoregressive and kernel spectral density estimators both

indicate $\delta=0$ at $\omega=0$ and $\delta=2$ at $\omega=0.08333$.

Comparison of Yule-Walker and Burg estimators of autoregressive coefficients and partial correlations. If the estimators are significantly different, we would conclude that the time series is long memory and an ARMA model is not applicable.

<u>Index</u>	<u>Autoregressive coefficients</u>		<u>Partial Correlations</u>	
	<u>Yule-Walker</u>	<u>Burg</u>	<u>Yule-Walker</u>	<u>Burg</u>
1	-.553	-.388	.818	.825
2	.020	-.093	-.634	-.661
3	.044	.023	-.497	-.576
4	.162	.117	-.454	-.491
5	.179	.222	-.332	-.399
6	.095	.129	-.192	-.221
7	.120	.170	-.153	-.175
8	.135	.144	-.047	-.048
9	-.162	-.255	.162	.255

Our current experience leads us to believe that the above estimators are just barely "significantly different." However the spectral densities seem to yield similar results.

Comparison of spectral estimators. The Burg AR spectral estimator is strongly peaked with peak at $\omega=0.0833$ (period 12). The spectral distribution rises from 0.03 to 0.96 over the interval (.076, .090) corresponding to periods (13.09, 11.08). The sample spectral distribution rises from .05 to .92, while the Yule-Walker autoregressive spectral distribution rises from .06 to .93.

ARMA Subset Regression Based on Estimating $MA(\infty)$. Our algorithm yields the canonical models $ARMA(1,2;12)$ and $AR(1,12,13)$. That an ARMA model should not be fit to the time series of NYC monthly temperatures is indicated by the lack of fit of the ARMA spectral distributions to the sample spectral distribution, since the former rise from .16 to .85 for cepstral-based $MA(\infty)$, and from .15 to .81 for Burg-based $MA(\infty)$, over the frequency interval (0.076, 0.090). We have not investigated whether the ARMA models identified would fit better if their parameters were estimated more efficiently than they are by our subset regression algorithm.

Conclusion. A model for $Y(t)$ which has maximum insight is: $Y(t) = S(t) + Z(t)$, where $S(t)$ is a function with period 12 [initially estimated by the monthly means], and $Z(t)$ is a stationary time series. The spectrum of interest here would seem to be that of $Z(t)$. However if one insists on a spectral density estimator for $Y(t) - \bar{Y}$, a satisfactory answer may be the autoregressive spectral density estimator of order 9 with coefficients computed by a Burg (or least squares) algorithm rather than by the Yule-Walker equations. Since AR order determining criteria do not apply for this model, the question is open if one should not base the AR spectral estimator on an $AR(13)$.

7. Summary of quantile spectral analysis

Given a sample $Y(t)$, $t=1,2,\dots,T$, quantile spectral analysis first forms the standardized time series $\{Y(t) - \text{median}\} \div \{\text{twice interquartile range}\}$ for which one computes the sample spectral density (periodogram), sample correlations, sample partial correlations, sample cepstral pseudo-correlations (and even sample inverse-correlations). The output we propose that one examine to identify time series memory and spectral density estimator is as follows:

(I) $\tilde{I}\tilde{Q}(u)$ and $\tilde{D}(u)$ plots of the original data (to identify its probability distribution), sample spectral density, and sample correlations.

(II) order determining criterion functions AIC and CAT; Yule-Walker estimators of autoregressive coefficients for best and second best AR orders; Burg estimators of autoregressive coefficients for the maximum of the best and second best AR orders.

(III) Diverse Spectral density (and corresponding Spectral distribution function) estimators computed by the following methods: (a) sample spectral density, (b) AR spectral density of best order with Yule-Walker computed coefficients; (c) AR spectral density with least squares (Burg algorithm) coefficients (d) ARMA Spectral density estimators with coefficients determined by subset regression, based on an $MA(\infty)$ representation computed from an approximating AR scheme, (e) ARMA spectral density estimators

with coefficients determined by subset regression, based on an $MA(\infty)$ representation formed from sample cepstral pseudo-correlations. Each of these methods also yields estimators of σ_{∞}^2 .

(IV) Each estimated spectral density is used to compute estimators δ_k of the index δ of regular variation of $f(\omega)$ at $\omega=0$ and a specified seasonal frequency. [a formula for δ_k is given below].

(V) An estimated spectral density is formed called the local quantile spectral estimator; it is based on the median and quartiles of the set of values of the sample spectral density in a specified neighborhood of an equi-spaced grid of frequencies.

The approach to time series model identification outlined in this paper can be considered exploratory data analysis since the diverse criterion functions utilized 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 theoretically the distribution of the criteria for various time series models. The ultimate validity of this approach (and refinements of its reasoning process) can only be accomplished by a series of examples of important practical applications.

Among the important questions for further research is more theory concerning the index δ of regular variation of a

spectral density $f(\omega)$ at a frequency ω_0 , defined by the representation

$$f(\omega) = (\omega - \omega_0)^{-\delta} L(\omega - \omega_0)$$

where $L(x)$ is slowly varying as x tends to 0. No and short memory time series have $\delta=0$ at all frequencies. Long memory time series have $\delta \neq 0$ at some frequency. To estimate δ from a consistent (windowed or AR) estimator $\hat{f}(\frac{k}{n})$ of the spectral density at a grid of equi-spaced frequencies, we choose m so that $m/n = \omega_0$ and form a sequence

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

One conjectures that if n and k are integers tending to ∞ in such a way that k/n tends to 0, then $\delta = \lim \delta_k$.

A value of $\delta=2$ indicates a sharp peak in the spectral density, that differencing once may be justified, or that a periodic signal should be fitted. A value of $\delta=-.2$ indicates a sharp trough in the spectral density which may be the result of over-differencing. The convergence of δ_k to δ is very slow, and we currently use the shape of the curve δ_k rather than any of its individual values as the evidence for interpretation.

REFERENCES

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

END

**DATE
FILMED**

10 — 83

DTIC