THE ROLE OF SPECTRAL ANALYSIS IN
TIME SERIES ANALYSIS

BY
EMANUEL PARZEN

TECHNICAL REPORT NO. 2
July 12, 1965

PREPARED UNDER CONTRACT N00014-225(80)
(NR-042-234)
FOR
OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ARCHIVE COPY
THE ROLE OF SPECTRAL ANALYSIS
IN TIME SERIES ANALYSIS

by

Emanuel Parzen

TECHNICAL REPORT NO. 2
July 12, 1965

PREPARED UNDER CONTRACT Nonr-225(80)
(NR-042-234)

FOR

OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
The Role of Spectral Analysis in Time Series Analysis

by

Emanuel Parzen
Stanford University

1. Introduction

Statistical spectral analysis has several roles in time series analysis: (i) estimation; (ii) hypothesis testing and hypothesis suggesting; and (iii) description and reduction of data.

In any field where the properties of the phenomenon being studied can be characterized in terms of its behavior in the frequency domain one needs to estimate spectral density functions and other spectral characteristics associated with stationary multiple time series.

However, spectral analytic techniques seem to provide also a means of testing the fit of various models (the goodness-of-fit of a model can be discussed using sample spectra of the residuals from the fitted model) and suggesting possible models to fit (explanatory "variables" or "mechanisms" to be fitted to a time series are often suggested by sample spectra). As stated so lucidly by Herman Wold (1947), "empirical time series present such a host of widely different patterns that the hypotheses about their structure cannot adequately be brought together into a single parameter system." Consequently, an analysis of a time series is not accomplished by adopting a single

Prepared with the partial support of the Office of Naval Research. Reproduction is permitted for any purpose of the United States Government. To be presented at the International Statistical Institute, September, 1965.
model, the parameters of which are estimated. Rather, it is best carried out by a process of increasing insight from successive analyses.

To a sample \( \{X(t), t = 1, 2, ..., T\} \) of a time series, one can associate a function, called the sample spectral density function or periodogram, defined by

\[
f_T(\omega) = \frac{1}{2\pi T} \sum_{t=1}^{T} e^{-i\omega t} X(t)^2, \quad -\pi \leq \omega \leq \pi.
\]

The periodogram was introduced by Schuster to estimate the frequencies of strict periodicities in a time series satisfying the following assumptions: (i) it is not evolving but is oscillating about a constant level; (ii) it may be regarded as composed of a number of "strict" periodicities plus purely random fluctuations.

Since the model of strict periodicities plus random noise seems to occur rarely in practice, the Schuster periodogram often discovered spurious cycles. To remedy this, the notion of "disturbed" periodicity was introduced by means of autoregressive models and moving average models. The correlogram came to the fore, and periodogram analysis fell into disfavor, if not disrepute.

Autoregressive and moving average models are (under some additional assumptions) special cases of stationary time series. The problems of finding the order of a finite parameter scheme such as autoregressive and moving average models led time series analysts to adopt a "non-parametric" approach and first estimate the spectral density function of the observed time series under the assumption that it was a stationary time series with no strict periodicities. Techniques of spectral analysis...
(using smoothed periodograms) came back into favor, when interpreted as estimates of the spectral density function of an underlying stationary time series.

But there is more to life than stationary time series with continuous spectra. Consequently, statisticians added the possibility of strict periodicities back to the model. Thus was born the so-called problem of mixed spectra (see Hext (1966) for a current survey).

Finally, the assumption that the observed time series is trend free is unnatural. If one adds trend to a stationary time series, one has a time series which can be regarded as derived from a stationary time series by a filtering process. A theory of Fourier analysis can be developed for such time series, and one might seek to estimate this spectrum. In our approach to empirical time series analysis, [see Parzen (1966)] the emphasis is on the use of spectra defined from samples rather than from populations or ensembles. Given an observed time series of finite length, or a time series derived from it, one defines various "sample spectral functions" such as windowed sample spectral density functions and distribution functions. Their properties can be determined for each possible model one desires to consider for the observed time series. Consequently, they can be used to form estimates of the parameters characterizing the model. Further, they can be used to determine an appropriate model by comparing the actual appearance of these spectral functions with their expected appearance under the various models; that model for which the correspondence is closest is considered the most likely.

Our aim in this paper is: (1) to summarize the basic formulas
employed in the empirical spectral analysis of a single time series, and
(2) to show their applicability to the problem of analyzing and
synthesizing "adaptive predictors" for time series. The empirical
spectral analysis of multiple time series is discussed in Parzen (1965).

Other uses of spectral analysis are described in the excellent
survey paper of Jenkins (1965).

2. Sample convolution function

In order to define windowed sample spectral density functions (or
smoothed periodograms) it is convenient to first introduce the sample
convolution function, denoted $R_t(\cdot)$, of an observed sample

\[ X(t), t = 1, 2, \ldots, T \]

\[ R_t(v) = \frac{1}{T} \sum_{t=1}^{T-v} X(t) X(t+v), \quad v = 0, 1, \ldots, T - 1, \]

(1)

\[ = R_t(-v), \quad v = -1, \ldots, -(T-1), \]

\[ = 0, \quad \text{otherwise}. \]

The terminology "sample convolution function" is not standard but
is introduced in this paper in order to reserve for other purposes the
terms "sample correlation function" or "autocorrelation function" which
are used by other authors. In the case that the time series of which
\[ X(t), t = 1, 2, \ldots, T \]
is a sample is known to have zero means and to be covariance stationary with covariance function $R(\cdot)$ then $R_t(v)$ provides a possible estimate of $R(v)$. Because of this, in previous
writings [see Parzen (1964 a), (1964 b)] the author has called the
function $R_t(\cdot)$ the sample covariance function. However, it is our
belief that the computation of $R_t(\cdot)$ is of great value even for time
series which are not necessarily covariance stationary. It seems best therefore when introducing this function for the first time to give it a name which indicates its data-handling, rather than statistical, character; such a name is "sample convolution function." Similarly, the sample correlation function $\rho_T(v)$, defined below, is from a statistical point of view, an estimate of the true correlation function $\rho(v)$ of a covariance stationary time series with zero means, while from a data-handling point of view, it is just the convolution function multiplied by a scale factor so as to have value 1 at $v = 0$.

The relations that exist between the sample convolution function and sample spectral density function of an observed time series are the same as those that exist between the covariance function and spectral density function of a covariance stationary time series. In particular, we note the following facts.

The sample convolution function $R_T(v)$ and the sample spectral density function $f_T(\omega)$ are both even functions of their arguments and are a Fourier transform pair:

$$R_T(v) = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{\cos(\omega v)}{\cos(\omega)} f_T(\omega) \, d\omega = 2 \int_0^\pi \cos \omega f_T(\omega) \, d\omega,$$

$$f_T(\omega) = \frac{1}{2\pi} R_T(\omega) + \frac{1}{\pi} \sum_{v=1}^{T-1} \cos \omega R_T(v).$$

The sample distribution function. Given an observed time series \{X(t), t = 1, 2, ..., T\}, the sample distribution function $F_T(\omega)$ is a function of $\omega$ in the interval $0 \leq \omega \leq \pi$ defined by
Conversely, \( R_T(v) \) is the Fourier-Stieltjes transform of \( F_T(\omega) \):

\[
R_T(v) = \int_{0}^{\pi} \cos v\omega \, dF_T(\omega)
\]

The spectral distribution function is a monotone increasing function of \( \omega \). Consequently it fluctuates much less than the sample spectral density function \( f_T(\omega) \). This is both a virtue and a vice. Certain real effects which it is the aim of the investigation to discern will show up most clearly in the spectral density function whereas they may be overlooked in the spectral distribution function. On the other hand, certain spurious effects may appear to show up in the spectral density function which on the basis of the spectral distribution function may be rejected as pure fluctuation.

Sample correlation function. The sample correlation function, denoted \( \rho_T(\cdot) \), is defined by

\[
\rho_T(v) = \frac{R_T(v)}{R_T(0)}
\]

In words, \( \rho_T(v) \) is the sample convolution function normalized to have value 1 at \( v = 0 \).

Normalized sample spectral density function. For ease of comparing
the sample spectral density functions arising from different time series, it seems best to compute and plot normalized versions of these functions. Since

\[ R_T(\omega) = \int_{-\pi}^{\pi} f_T^T(\omega) \, d\omega , \]

the natural normalization of \( f_T^T(\omega) \) is

\[ f_T^*(\omega) = \frac{f_T^T(\omega)}{R_T(0)} \]

which has the property that its integral form \( -\pi \) to \( \pi \) equals 1. We call \( f_T^*(\omega) \) the normalized spectral density function; note that it is also the spectral density function of the sample correlation function,

\[ R_T(v) = \int_{-\pi}^{\pi} e^{i\omega v} f_T^*(\omega) \, d\omega . \]

3. Windowed sample spectral density and distribution functions

The windowed sample spectral density function, denoted \( f_{T,M}^T(\omega) \), is defined by (for \( -\pi \leq \omega \leq \pi \))

\[ f_{T,M}^T(\omega) = \frac{1}{2\pi} \sum_{|v| \leq M} \cos \omega k(v) R_T(v) \]

where \( R_T(\cdot) \) is the sample convolution function.

The windowed normalized sample spectral density function, denoted \( f_{T,M}^*(\omega) \), is defined by (for \( -\pi \leq \omega \leq \pi \))

\[ f_{T,M}^*(\omega) = \frac{1}{2\pi} \sum_{|v| \leq M} \cos \omega k(v) \rho_T(v) . \]
where \( \rho_T(\cdot) \) is the sample correlation function.

The windowed normalized sample spectral distribution function, denoted \( \overline{F}_{T,M}(\omega) \), is defined by (for \( 0 \leq \omega \leq \pi \))

\[
\overline{F}_{T,M}(\omega) = 2 \int_0^\omega \overline{F}_{T,M}(\omega') \, d\omega',
\]

\[= \frac{\omega}{\pi} + \frac{2}{\pi} \sum_{v=1}^{M} \frac{\sin v\omega}{v} k\left(\frac{v}{M}\right) \rho_T(v) \tag{3}\]

**Lag Windows.** The function \( k(\cdot) \) is known as the lag window of the windowed spectrum. In our work we use mainly the following lag window

\[
k(u) = 1 - 6u^2 + 6|u|^3, \quad |u| \leq 0.5
\]

\[
= 2(1-|u|)^3, \quad 0.5 \leq |u| \leq 1.0
\]

\[
= 0, \quad |u| \geq 1.
\]

A kernel widely used in existing spectral analysis programs is one suggested by Tukey (see Blackman and Tukey (1958), p. 14):

\[
k(u) = \frac{1}{2} (1 + \cos \pi u), \quad |u| < 1,
\]

\[
= 0, \quad \text{otherwise}.
\]

This lag window is not used in our work because the corresponding windowed spectrum is not necessarily non-negative (and the corresponding estimates of coherence are not necessarily between 0 and 1).

**Truncation Points.** The integer \( M(<T) \) is called the truncation
point of the windowed spectrum since it represents the number of sample
correlations of the $T$ available actually used in computing the
spectrum. It is wise to choose several truncation points in practice.
In our computations, we usually choose three truncation points
$M_1, M_2, M_3$ as percentages of $T$:

$$5\% \leq \frac{M_1}{T} \leq 10\%, \quad 10\% \leq \frac{M_2}{T} \leq 25\%, \quad 25\% \leq \frac{M_3}{T} \leq 75\%.$$ 

An alternative rule is:

$$5\% T \leq M_1 \leq 10\% T, \quad 2M_1 \leq M_2 \leq 3M_1, \quad 2M_2 \leq M_3 \leq 3M_2.$$ 

**Spectral Computation Number.** There is a third choice to be made in
forming the estimate $f_{T,M}(\omega)$, and this is the number of points on the
interval $0$ to $\pi$ at which it will be computed. We adopt the
attitude that $f_{T,M}(\omega)$ should be computed for equispaced frequencies

$$\omega = 0, \frac{\pi}{Q}, \frac{2\pi}{Q}, \ldots, \pi$$

where $Q$ is an integer to be chosen. We call $Q$ the spectral computa-
tion number.

In the past $Q$ has frequently been chosen to be equal to the
truncation point $M$. One can prove a sampling theorem to the effect
that the estimated spectrum (which is a function of $\omega$, measured in
cycles per unit of observation time, in the interval $0 \leq \omega \leq 0.5$)
can be recovered from its value at $M$ equally spaced points. However,
this recovery cannot necessarily be done by linear interpolation. If
the graph of the estimated spectrum is to be obtained by merely drawing
line segments connecting the computed values, one needs to compute the 
spectrum at $Q$ equi-spaced frequencies, where $Q$ should be at least 
$2M$ and perhaps should be $4M$ (note: further research is needed on 
this point).

If one uses $3$ truncation points $M_1 < M_2 < M_3$, it has seemed 
reasonable to me to compute each spectrum at $Q = M_3$ points. However, 
one should choose $Q$ (approximately equal to $M_3$) such that the 
frequencies which are multiples of $\pi/Q$ are of physical interest. For 
economic time series of monthly data $w_\alpha$, usually choose $Q$ to be a 
multiple of $12$.

**Spectral Window.** The *spectral window* of the windowed spectrum 
defined by (2) is defined to be the function

$$K_M(\omega) = \frac{1}{2\pi} \sum_{|v| \leq M} e^{i\omega v} k(v)$$

(6)

For the lag window (4), it may be shown that

$$K_M(\omega) = \frac{3}{8\pi^3} \left[ \sin \left( \frac{M\omega}{4} \right) \right]^4 \left[ 1 - \frac{2}{3} \left( \sin \frac{\omega}{2} \right)^2 \right]$$

(7)

This is an even function which integrates to $1$, has maximum value

$$K_M(0) = \frac{3}{8\pi} M$$

(8)

and achieves its first zero at $\omega = 4\pi/M$. It is thus concentrated 
about $\omega = 0$ with a rectangular bandwidth $8\pi/3M$ in radians and 
$4/3M$ in cycles per unit time.

In order to understand the name "spectral window," we must first
note the relation that exists between the sample spectral density function $f_T(\omega)$ and the windowed spectrum $f_{T, M}(\omega)$:

\begin{equation}
(9) \quad f_{T, M}(\omega) = \int_{-\pi}^{\pi} K_M(\omega-\lambda) f_T(\lambda) \, d\lambda
\end{equation}

since

\begin{equation}
(10) \quad f_{T, M}(\omega) = \frac{1}{2\pi} \sum_{\nu \leq M} e^{-i\nu \omega} K_M(\nu) \int_{-\pi}^{\pi} e^{i\nu \lambda} f_T(\lambda) \, d\lambda
\end{equation}

Thus $f_{T, M}(\omega)$ is the convolution of $f_T(\omega)$ and $K_M(\omega)$. In other words, $f_{T, M}(\omega)$ is an averaging over the values of $f_T(\omega)$ when it is viewed through a window (or channel) of variable transmission properties given by $K_M(\omega)$.

A useful approximation to $K_M(\omega)$ can be obtained by introducing the Fourier transform

\begin{equation}
(11) \quad K(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega u} k(u) \, du
\end{equation}

which we call the spectral window generator. It may be shown that approximately

\begin{equation}
(12) \quad K_M(\omega) = M K(M\omega)
\end{equation}

since exactly

\begin{equation}
(13) \quad K_M(\omega) = M \sum_{j=-\infty}^{\infty} K(M(\omega-2\pi j))
\end{equation}
From this basic formula we obtain the formula (7) for $K_M(\omega)$. For
the lag window (4), the spectral window generator may be shown to be

$$K(\omega) = \frac{3}{\pi^2} \left\{ \frac{\sin(\omega/4)}{\omega/4} \right\}^4$$

(14)

Then

$$MK(M\omega) = \frac{3}{\pi M^2} \left\{ \frac{\sin(M\omega/4)}{\omega/4} \right\}^4$$

(15)

approximates $K_M(\omega)$ given by (7).

Rather than giving a theoretical discussion of the properties of
windowed sample spectra we illustrate their use by analyzing a time
series which has been extensively discussed from the point of view of
forecasting. This is a monthly series of international airline passen-
ger bookings, 1949-1961; compare Brown (1963), p. 429 and Barnard
(1963).

It is to be emphasized that there is not a uniquely best way in
which spectral analytic ideas enter into time series analysis once one
drops the assumption that one is dealing with a stationary time series.
The attitudes to data analysis presented in this paper should be used
in conjunction with other attitudes such as computing time varying
spectra.

4. Analysis of an empirical time series

The monthly time series of international airline passengers has
the characteristic features of many social and economic time series; in
particular, there is an upward trend and a seasonal variation. In
In order to gain insight into the structure of a time series, we often seek to find the coefficients of the minimum mean square error linear predictor \( \hat{X}(t) \) of the time series \( X(t) \) of the form

\[
(1) \quad \hat{X}(t) = a_1 X(t-1) + \ldots + a_m X(t-m)
\]

There are three ways in which one can fit an autoregressive scheme to data: (i) one can specify the order \( m \) and estimate the coefficients \( a_i \) by solving the system of linear equations

\[
(2) \quad E[\hat{X}(t) X(t-i)] = E[X(t) X(t-i)], \quad i = 1, 2, \ldots, m;
\]

(ii) one can take the possible order \( m \) to be some large number (such as 50 months) but admit only those lags whose coefficients \( a_i \) are "significantly" different from zero; (iii) one can take the possible order \( m \) to be some large number but solve for the coefficients \( a_i \) in order of decreasing contribution to the residual sum of squares, and use only a specified number of coefficients. Applying these procedures to the airline passenger series we find the following results: (i) if one fits a 13th order scheme,
The coefficients are written in order of decreasing contribution to mean square prediction error. Next let us seek only the coefficients making a "significant" contribution to the residual sum of squares; one would then fit a first order predictor

\[ \hat{x}_1(t) = 0.985 \ x(t-1) + 0.248 \ x(t-11) - 0.346 \ x(t-13) - 0.134 \ x(t-6) + 0.190 \ x(t-12) + 0.068 \ x(t-2) - 0.080 \ x(t-3) + 0.050 \ x(t-4) - 0.007 \ x(t-8) + 0.019 \ x(t-5) - 0.008 \ x(t-7) + 0.012 \ x(t-10) - 0.011 \ x(t-9) \]

Finally let us arbitrarily choose to fit the best fitting 3 terms; one obtains the predictor

\[ \hat{x}_2(t) = 0.986 \ x(t-1) \]

\[ \hat{x}_3(t) = 0.957 \ x(t-1) + 0.257 \ x(t-11) - 0.224 \ x(t-13) \]
The means and variances of the original series and the residual series are as follows:

<table>
<thead>
<tr>
<th>Series</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X(t)$</td>
<td>280</td>
<td>$1.43 \times 10^1$</td>
</tr>
</tbody>
</table>

- $\epsilon_1(t) = \hat{X}(t) - X_1(t)$
  - Mean: 5.36
  - Variance: $5.13 \times 10^2$

- $\epsilon_2(t) = \hat{X}(t) - X_2(t)$
  - Mean: 5.39
  - Variance: $1.11 \times 10^3$

- $\epsilon_3(t) = \hat{X}(t) - X_3(t)$
  - Mean: 4.81
  - Variance: $5.28 \times 10^2$

Next one examines the spectra of the residual series $\epsilon(t) = X(t) - \hat{X}(t)$. The windowed sample spectral density function and spectral distribution function of $\epsilon_3(t)$ is plotted in the top half of figures 2 and 3, respectively; the trend in the original $X(t)$ series has been eliminated, but the seasonal peaks remain. The spectra of $\epsilon_1(t)$ and $\epsilon_2(t)$ are similar except that $\epsilon_2(t)$ has stronger seasonal peaks.

We next repeat the autoregressive model fitting procedures on the residual $\epsilon_2(t)$ and $\epsilon_3(t)$. We find (applying procedure 2 to $\epsilon_2(t)$):

\[
\hat{\epsilon}_2(t) = 0.831 \epsilon_2(t-12)
\]

while (applying procedure 3 to $\epsilon_3(t)$):

\[
\hat{\epsilon}_3(t) = 0.657 \epsilon_3(t-12) + 0.141 \epsilon_3(t-24) - 0.080 \epsilon_3(t-20)
\]

In order to interpret the properties of the residuals $\eta(t) = \epsilon(t) - \hat{\epsilon}(t)$ let us compare them with the forecasting errors.
given by Barnard (1963) in his comparison of the "adaptive forecasting" and Box-Jenkins method. Data for 1949 and 1950 were used to provide initial values, so that the forecast errors were given only for the 10 years, 1951-1960. For comparison we computed the spectra of the forecasting errors $\hat{\eta}_2(t)$ and $\hat{\eta}_3(t)$ over this 10-year period.

<table>
<thead>
<tr>
<th>Series of forecasting errors, 1951-1960</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive forecasting method</td>
<td>.28</td>
<td>1.78 $10^2$</td>
</tr>
<tr>
<td>Box-Jenkins method</td>
<td>-.36</td>
<td>1.83 $10^2$</td>
</tr>
<tr>
<td>$\eta_2(t) = \epsilon_2(t) - \hat{\epsilon}_2(t)$</td>
<td>1.52</td>
<td>2.05 $10^2$</td>
</tr>
<tr>
<td>$\eta_3(t) = \epsilon_3(t) - \hat{\epsilon}_3(t)$</td>
<td>1.76</td>
<td>1.73 $10^2$</td>
</tr>
</tbody>
</table>

The spectra of the forecasting errors arising from adaptive forecasting and the Box-Jenkins method are very different! They both are far from the spectrum of white noise, but the adaptive forecasting errors are predominantly low frequency while the Box-Jenkins forecasting errors are predominantly high frequency; their sample windowed spectral density functions and spectral distribution functions are plotted in figures 4 and 5, respectively, for a Parzen window and in figures 6 and 7 for a Tukey window.

In the bottom half of figures 2 and 3 we plot the sample spectrum of $\eta_3(t)$; it is essentially the spectrum of white noise.

The foregoing considerations lead to both a model and a forecasting formula for $X(t)$. Let $U_r$ be the $r$-th backward shift operator, $U_rX(t) = X(t-r)$, and let $I$ be the identity operator. Define
\[ P_1 = 0.957 U_1 + 0.257 U_{11} - 0.274 U_{13} , \]

\[ P_2 = 0.657 U_{12} - 0.080 U_{20} + 0.141 U_{24} . \]

Then there is a white-noise series \( \eta(t) \) such that

\[ (I-P_2) (I-P_1) X(t) = \eta(t) \]

Therefore a predictor of \( X(t) \) is given by

\[ \hat{X}(t) = (P_1 + P_2 - P_2 P_1) X(t) \]

More generally, let \( \hat{X}(t+v) \) denote the predictor of \( X(t+v) \) given values of the time series up to time \( t \). Then

\[ \hat{X}(t+v) = (P_1 + P_2 - P_2 P_1) \hat{X}(t+v) ; \]

note that \( \hat{X}(s) = X(s) \) if \( s \leq t \).

The aim of the foregoing discussion has been to show one important use of empirical spectral analysis; given an operator (such as \( I - P_1 \)) the properties of the time series \( (I - P_1) X(t) \) can be studied without regard to the procedure by which one formed the operator.
References


