Chapter 17
The SPECTRA Procedure

Chapter Table of Contents

Chapter 17
The SPECTRA Procedure

Overview

The SPECTRA procedure performs spectral and cross-spectral analysis of time series. You can use spectral analysis techniques to look for periodicities or cyclical patterns in data.

The SPECTRA procedure produces estimates of the spectral and cross-spectral densities of a multivariate time series. Estimates of the spectral and cross-spectral densities of a multivariate time series are produced using a finite Fourier transform to obtain periodograms and cross-periodograms. The periodogram ordinates are smoothed by a moving average to produce estimated spectral and cross-spectral densities. PROC SPECTRA can also test whether or not the data are white noise.

PROC SPECTRA uses the finite Fourier transform to decompose data series into a sum of sine and cosine waves of different amplitudes and wavelengths. The Fourier transform decomposition of the series x_t is

$$x_t = \frac{a_0}{2} + \sum_{k=1}^{m} [a_k cos(\omega_k t) + b_k sin(\omega_k t)]$$

where

t	is the time subscript, $t = 1, 2, \ldots, n$
x_t	are the data
n	is the number of observations in the time series
m	is the number of of frequencies in the Fourier decomposition: $m = \frac{n}{2}$ if n is even; $m = \frac{n-1}{2}$ if n is odd
a_0	is the mean term: $a_0 = 2\overline{x}$
a_k	are the cosine coefficients
b_k	are the sine coefficients
ω_k	are the Fourier frequencies: $\omega_k = \frac{2\pi k}{n}$

Functions of the Fourier coefficients a_k and b_k can be plotted against frequency or against wave length to form *periodograms*. The amplitude periodogram J_k is defined as follows:

$$J_k = \frac{n}{2}(a_k^2 + b_k^2)$$

Several definitions of the term periodogram are used in the spectral analysis literature. The following discussion refers to the J_k sequence as the periodogram.

The periodogram can be interpreted as the contribution of the *k*th harmonic ω_k to the total sum of squares, in an analysis of variance sense, for the decomposition of the process into two-degree-of-freedom components for each of the *m* frequencies. When *n* is even, $sin(\omega_{\frac{n}{2}})$ is zero, and thus the last periodogram value is a one-degree-of-freedom component.

The periodogram is a volatile and inconsistent estimator of the spectrum. The spectral density estimate is produced by smoothing the periodogram. Smoothing reduces the variance of the estimator but introduces a bias. The weight function used for the smoothing process, W(), often called the kernel or spectral window, is specified with the WEIGHTS statement. It is related to another weight function, *w*(), the lag window, that is used in other methods to taper the correlogram rather than to smooth the periodogram. Many specific weighting functions have been suggested in the literature (Fuller 1976, Jenkins and Watts 1968, Priestly 1981). Table 17.1 later in this chapter gives the formulas relevant when the WEIGHTS statement is used.

Letting *i* represent the imaginary unit $\sqrt{-1}$, the cross-periodogram is defined as follows:

$$J_k^{xy} = \frac{n}{2}(a_k^x a_k^y + b_k^x b_k^y) + i\frac{n}{2}(a_k^x b_k^y - b_k^x a_k^y)$$

The cross-spectral density estimate is produced by smoothing the cross-periodogram in the same way as the periodograms are smoothed using the spectral window specified by the WEIGHTS statement.

The SPECTRA procedure creates an output SAS data set whose variables contain values of the periodograms, cross-periodograms, estimates of spectral densities, and estimates of cross-spectral densities. The form of the output data set is described in the section "OUT= Data Set" later in this chapter.

Getting Started

To use the SPECTRA procedure, specify the input and output data sets and options for the analysis you want on the PROC SPECTRA statement, and list the variables to analyze in the VAR statement.

For example, to take the Fourier transform of a variable X in a data set A, use the following statements:

```
proc spectra data=a out=b coef;
   var x;
run;
```

This PROC SPECTRA step writes the Fourier coefficients a_k and b_k to the variables COS_01 and SIN_01 in the output data set B.

When a WEIGHTS statement is specified, the periodogram is smoothed by a weighted moving average to produce an estimate for the spectral density of the series. The following statements write a spectral density estimate for X to the variable S_01 in the output data set B.

```
proc spectra data=a out=b s;
   var x;
   weights 1 2 3 4 3 2 1;
run;
```

When the VAR statement specifies more than one variable, you can perform cross-spectral analysis by specifying the CROSS option. The CROSS option by itself produces the cross-periodograms. For example, the following statements write the real and imaginary parts of the cross-periodogram of X and Y to the variable RP_01_02 and IP_01_02 in the output data set B.

```
proc spectra data=a out=b cross;
   var x y;
run;
```

To produce cross-spectral density estimates, combine the CROSS option and the S option. The cross-periodogram is smoothed using the weights specified by the WEIGHTS statement in the same way as the spectral density. The squared coherency and phase estimates of the cross-spectrum are computed when the K and PH options are used.

The following example computes cross-spectral density estimates for the variables X and Y.

```
proc spectra data=a out=b cross s;
   var x y;
   weights 1 2 3 4 3 2 1;
run;
```

The real part and imaginary part of the cross-spectral density estimates are written to the variable CS_01_02 and QS_01_02, respectively.

Syntax

The following statements are used with the SPECTRA procedure.

PROC SPECTRA *options*;
BY *variables*;
VAR *variables*;
WEIGHTS *constants*;

Functional Summary

The statements and options controlling the SPECTRA procedure are summarized in the following table.

Description	Statement	Option
Statements		
specify BY-group processing	BY	
specify the variables to be analyzed	VAR	
specify weights for spectral density estimates	WEIGHTS	
Data Set Options		
specify the input data set	PROC SPECTRA	DATA=
specify the output data set	PROC SPECTRA	OUT=
Output Control Options		
output the amplitudes of the cross-spectrum	PROC SPECTRA	A
output the Fourier coefficients	PROC SPECTRA	COEF
output the periodogram	PROC SPECTRA	P
output the spectral density estimates	PROC SPECTRA	S
output cross-spectral analysis results	PROC SPECTRA	CROSS
output squared coherency of the cross-spectrum	PROC SPECTRA	K
output the phase of the cross-spectrum	PROC SPECTRA	PH
Smoothing Options		
specify the Bartlett kernel	WEIGHTS	BART
specify the Parzen kernel	WEIGHTS	PARZEN
specify the Quadratic Spectral kernel	WEIGHTS	QS
specify the Tukey-Hanning kernel	WEIGHTS	TUKEY
specify the Truncated kernel	WEIGHTS	TRUNCAT

Description	Statement	Option
Other Options		
subtract the series mean	PROC SPECTRA	ADJMEAN
specify an alternate quadrature spectrum estimate	PROC SPECTRA	ALTW
request tests for white noise	PROC SPECTRA	WHITETEST

PROC SPECTRA Statement

PROC SPECTRA *options;*

The following options can be used in the PROC SPECTRA statement.

A
outputs the amplitude variables (A_*nn*_*mm*) of the cross-spectrum.

ADJMEAN
CENTER
subtracts the series mean before performing the Fourier decomposition. This sets the first periodogram ordinate to 0 rather than 2*n* times the squared mean. This option is commonly used when the periodograms are to be plotted to prevent a large first periodogram ordinate from distorting the scale of the plot.

ALTW
specifies that the quadrature spectrum estimate is computed at the boundaries in the same way as the spectral density estimate and the cospectrum estimate are computed.

COEF
outputs the Fourier cosine and sine coefficients of each series, in addition to the periodogram.

CROSS
is used with the P and S options to output cross-periodograms and cross-spectral densities.

DATA= *SAS-data-set*
names the SAS data set containing the input data. If the DATA= option is omitted, the most recently created SAS data set is used.

K
outputs the squared coherency variables (K_*nn*_*mm*) of the cross-spectrum. The K_*nn*_*mm* variables are identically 1 unless weights are given in the WEIGHTS statement and the S option is specified.

OUT= *SAS-data-set*
names the output data set created by PROC SPECTRA to store the results. If the OUT= option is omitted, the output data set is named using the DATA*n* convention.

P
outputs the periodogram variables. The variables are named P_*nn*, where *nn* is an index of the original variable with which the periodogram variable is associated. When both the P and CROSS options are specified, the cross-periodogram variables RP_*nn_mm* and IP_*nn_mm* are also output.

PH
outputs the phase variables (PH_*nn_mm*) of the cross-spectrum.

S
outputs the spectral density estimates. The variables are named S_*nn*, where *nn* is an index of the original variable with which the estimate variable is associated. When both the S and CROSS options are specified, the cross-spectral variables CS_*nn_mm* and QS_*nn_mm* are also output.

WHITETEST
prints a test of the hypothesis that the series are white noise. See "White Noise Test" later in this chapter for details.

Note that the CROSS, A, K, and PH options are only meaningful if more than one variable is listed in the VAR statement.

BY Statement

BY *variables;*

A BY statement can be used with PROC SPECTRA to obtain separate analyses for groups of observations defined by the BY variables.

VAR Statement

VAR *variables;*

The VAR statement specifies one or more numeric variables containing the time series to analyze. The order of the variables in the VAR statement list determines the index, *nn*, used to name the output variables. The VAR statement is required.

WEIGHTS Statement

WEIGHTS *constant-specification* | *kernel-specification;*

The WEIGHTS statement specifies the relative weights used in the moving average applied to the periodogram ordinates to form the spectral density estimates. A WEIGHTS statement must be used to produce smoothed spectral density estimates. If the WEIGHTS statement is not used, only the periodogram is produced.

Using Constant Specifications

Any number of weighting constants can be specified. The constants should be positive and symmetric about the middle weight. The middle constant, (or the constant to the right of the middle if an even number of weight constants are specified), is the relative weight of the current periodogram ordinate. The constant immediately following the middle one is the relative weight of the next periodogram ordinate, and so on. The actual weights used in the smoothing process are the weights specified in the WEIGHTS statement scaled so that they sum to $\frac{1}{4\pi}$.

The moving average reflects at each end of the periodogram. The first periodogram ordinate is not used; the second periodogram ordinate is used in its place.

For example, a simple triangular weighting can be specified using the following WEIGHTS statement:

```
weights 1 2 3 2 1;
```

Using Kernel Specifications

You can specify five different kernels in the WEIGHTS statement. The syntax for the statement is

WEIGHTS [PARZEN][BART][TUKEY][TRUNCAT][QS] [c e];

where $c >= 0$ and $e >= 0$ are used to compute the bandwidth parameter as

$$\mathrm{l(q)} = \mathrm{cq}^{\mathrm{e}}$$

and q is the number of periodogram ordinates +1:

$$q = \mathrm{floor(n/2)} + 1$$

To specify the bandwidth explicitly, set $c =$ to the desired bandwidth and $e = 0$.

For example, a Parzen kernel can be specified using the following WEIGHTS statement:

```
weights parzen 0.5 0;
```

For details, see the “Kernels” section on page 982, later in this chapter.

Details

Input Data

Observations in the data set analyzed by the SPECTRA procedure should form ordered, equally spaced time series. No more than 99 variables can be included in the analysis.

Data are often de-trended before analysis by the SPECTRA procedure. This can be done by using the residuals output by a SAS regression procedure. Optionally, the data can be centered using the ADJMEAN option in the PROC SPECTRA statement, since the zero periodogram ordinate corresponding to the mean is of little interest from the point of view of spectral analysis.

Missing Values

Missing values are not supported by the SPECTRA procedure. If the SPECTRA procedure encounters a missing value for any variable listed in the VAR statement, it prints an error message and stops.

Computational Method

If the number of observations *n* factors into prime integers that are less than or equal to 23, and the product of the square-free factors of *n* is less than 210, then PROC SPECTRA uses the Fast Fourier Transform developed by Cooley and Tukey and implemented by Singleton (1969). If *n* cannot be factored in this way, then PROC SPECTRA uses a Chirp-Z algorithm similar to that proposed by Monro and Branch (1976). To reduce memory requirements, when *n* is small the Fourier coefficients are computed directly using the defining formulas.

Kernels

Kernels are used to smooth the periodogram by using a weighted moving average of nearby points. A smoothed periodogram is defined by the following equation.

$$\hat{J}_i(\mathrm{l(q)}) = \sum_{\tau=-\mathrm{l(q)}}^{\mathrm{l(q)}} \mathrm{w}\left(\frac{\tau}{\mathrm{l(q)}}\right) \tilde{\mathrm{J}}_{\mathrm{i}+\tau}$$

where $\mathrm{w(x)}$ is the kernel or weight function. At the endpoints, the moving average is computed cyclically; that is,

$$\tilde{J}_{i+\tau} = \begin{cases} J_{i+\tau} & \text{0 <= i+}\tau\text{ <= q} \\ J_{-(i+\tau)} & \text{i+}\tau\text{ < 0} \\ J_{q-(i+\tau)} & \text{i+}\tau\text{ > q} \end{cases}$$

The SPECTRA procedure supports the following kernels. They are listed with their default bandwidth functions.

Bartlett: KERNEL BART

$$\begin{aligned} \mathrm{w(x)} &= \begin{cases} 1-|x| & |x|\leq 1 \\ 0 & \text{otherwise} \end{cases} \\ \mathrm{l(q)} &= \frac{1}{2}q^{1/3} \end{aligned}$$

Parzen: KERNEL PARZEN

$$\begin{aligned} \mathrm{w(x)} &= \begin{cases} 1-6|x|^2+6|x|^3 & 0\leq|x|\leq\frac{1}{2} \\ 2(1-|x|)^3 & \frac{1}{2}\leq|x|\leq 1 \\ 0 & \text{otherwise} \end{cases} \\ \mathrm{l(q)} &= q^{1/5} \end{aligned}$$

Quadratic Spectral: KERNEL QS

$$\begin{aligned} \mathrm{w(x)} &= \frac{25}{12\pi^2x^2}\left(\frac{sin(6\pi x/5)}{6\pi x/5}-cos(6\pi x/5)\right) \\ \mathrm{l(q)} &= \frac{1}{2}q^{1/5} \end{aligned}$$

Tukey-Hanning: KERNEL TUKEY

$$\begin{aligned} \mathrm{w(x)} &= \begin{cases} (1+cos(\pi x))/2 & |x|\leq 1 \\ 0 & \text{otherwise} \end{cases} \\ \mathrm{l(q)} &= \frac{2}{3}q^{1/5} \end{aligned}$$

Truncated: KERNEL TRUNCAT

$$\begin{aligned} \mathrm{w(x)} &= \begin{cases} 1 & |x|\leq 1 \\ 0 & \text{otherwise} \end{cases} \\ \mathrm{l(q)} &= \frac{1}{4}q^{1/5} \end{aligned}$$

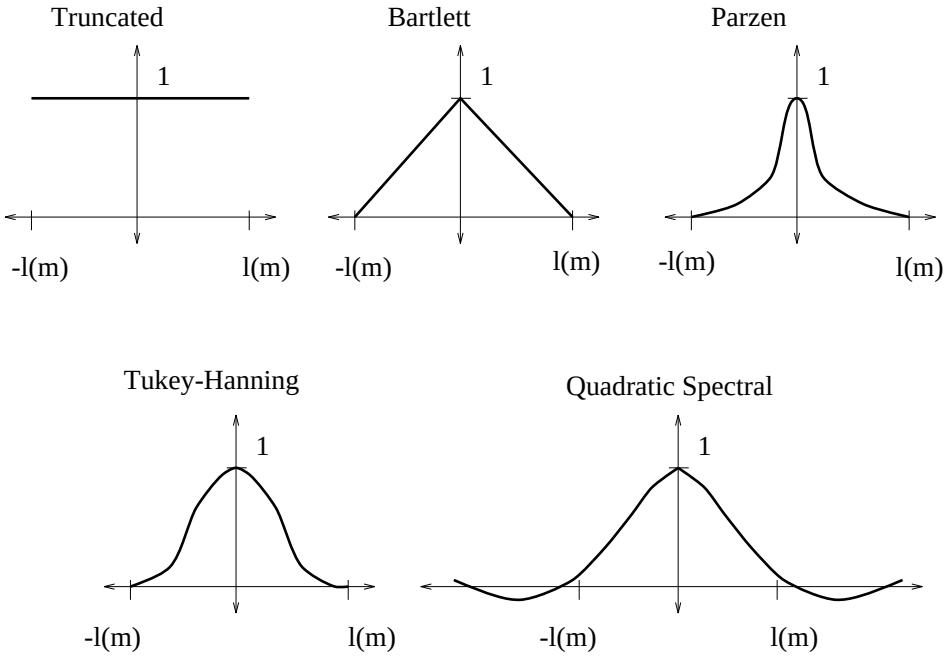


Figure 17.1. Kernels for Smoothing

Refer to Andrews (1991) for details on the properties of these kernels.

White Noise Test

PROC SPECTRA prints two test statistics for white noise when the WHITETEST option is specified: Fisher's Kappa (Davis 1941, Fuller 1976) and Bartlett's Kolmogorov-Smirnov statistic (Bartlett 1966, Fuller 1976, Durbin 1967).

If the time series is a sequence of independent random variables with mean 0 and variance σ^2, then the periodogram, J_k, will have the same expected value for all k. For a time series with nonzero autocorrelation, each ordinate of the periodogram, J_k, will have different expected values. The Fisher's Kappa statistic tests whether the largest J_k can be considered different from the mean of the J_k. Critical values for the Fisher's Kappa test can be found in Fuller 1976 and *SAS/ETS Software: Applications Guide 1*.

The Kolmogorov-Smirnov statistic reported by PROC SPECTRA has the same asymtotic distribution as Bartlett's test (Durbin 1967). The Kolmogorov-Smirnov statistic compares the normalized cumulative periodogram with the cumulative distribution function of a uniform(0,1) random variable. The normalized cumulative periodogram, F_j, of the series is

$$F_j = \frac{\sum_{k=1}^{j} J_k}{\sum_{k=1}^{m} J_k}, j = 1, 2 \ldots, m-1$$

where $m = \frac{n}{2}$ if n is even or $m = \frac{n-1}{2}$ if n is odd. The test statistic is the maximum absolute difference of the normalized cumulative periodogram and the uniform cu-

mulative distribution function. For $m-1$ greater than 100, if Bartlett's Kolmogorov-Smirnov statistic exceeds the critical value

$$\frac{a}{\sqrt{m-1}}$$

where $a = 1.36$ or $a = 1.63$ corresponding to 5% or 1% significance levels respectively, then reject the null hypothesis that the series represents white noise. Critical values for $m - 1 < 100$ can be found in a table of significance points of the Kolmogorov-Smirnov statistics with sample size $m - 1$ (Miller 1956, Owen 1962).

Transforming Frequencies

The variable FREQ in the data set created by the SPECTRA procedure ranges from 0 to π. Sometimes it is preferable to express frequencies in cycles per observation period, which is equal to $\frac{2}{\pi}$FREQ.

To express frequencies in cycles per unit time (for example, in cycles per year), multiply FREQ by $\frac{d}{2\pi}$, where *d* is the number of observations per unit of time. For example, for monthly data, if the desired time unit is years then *d* is 12. The period of the cycle is $\frac{2\pi}{d \times \mathrm{FREQ}}$, which ranges from $\frac{2}{d}$ to infinity.

OUT= Data Set

The OUT= data set contains $\frac{n}{2} + 1$ observations, if *n* is even, or $\frac{n+1}{2}$ observations, if *n* is odd, where *n* is the number of observations in the time series.

The variables in the new data set are named according to the following conventions. Each variable to be analyzed is associated with an index. The first variable listed in the VAR statement is indexed as 01, the second variable as 02, and so on. Output variables are named by combining indexes with prefixes. The prefix always identifies the nature of the new variable, and the indices identify the original variables from which the statistics were obtained.

Variables containing spectral analysis results have names consisting of a prefix, an underscore, and the index of the variable analyzed. For example, the variable S_01 contains spectral density estimates for the first variable in the VAR statement. Variables containing cross-spectral analysis results have names consisting of a prefix, an underscore, the index of the first variable, another underscore, and the index of the second variable. For example, the variable A_01_02 contains the amplitude of the cross-spectral density estimate for the first and second variables in the VAR statement.

Table 17.1 shows the formulas and naming conventions used for the variables in the OUT= data set. Let X be variable number *nn* in the VAR statement list and let Y be variable number *mm* in the VAR statement list. Table 17.1 shows the output variables containing the results of the spectral and cross-spectral analysis of X and Y.

In Table 17.1 the following notation is used. Let W_j be the vector of $2p + 1$ smoothing weights given by the WEIGHTS statement, normalized to sum to $\frac{1}{4\pi}$. The sub-

script of W_j runs from W_{-p} to W_p, so that W_0 is the middle weight in the WEIGHTS statement list. Let $\omega_k = \frac{2\pi k}{n}$, where $k = 0, 1, \ldots, \mathrm{floor}(\frac{n}{2})$.

Table 17.1. Variables Created by PROC SPECTRA

Variable	Description
FREQ	frequency in radians from 0 to π (Note: Cycles per observation is $\frac{\mathrm{FREQ}}{2\pi}$.)
PERIOD	period or wavelength: $\frac{2\pi}{\mathrm{FREQ}}$ (Note: PERIOD is missing for FREQ=0.)
COS_X COS_WAVE	cosine transform of X: $a_k^x = \frac{2}{n}\sum_{t=1}^{n} X_t \cos(\omega_k(t-1))$
SIN_X SIN_WAVE	sine transform of X: $b_k^x = \frac{2}{n}\sum_{t=1}^{n} X_t \sin(\omega_k(t-1))$
P_*nn*	periodogram of X: $J_k^x = \frac{n}{2}[(a_k^x)^2 + (b_k^x)^2]$
S_*nn*	spectral density estimate of X: $F_k^x = \sum_{j=-p}^{p} W_j J_{k+j}^x$ (except across endpoints)
RP_*nn_mm*	real part of cross-periodogram X and Y: $\mathrm{real}(\mathrm{J_k^{xy}}) = \frac{\mathrm{n}}{2}(\mathrm{a_k^x a_k^y} + \mathrm{b_k^x b_k^y})$
IP_*nn_mm*	imaginary part of cross-periodogram of X and Y: $\mathrm{imag}(\mathrm{J_k^{xy}}) = \frac{\mathrm{n}}{2}(\mathrm{a_k^x b_k^y} - \mathrm{b_k^x a_k^y})$
CS_*nn_mm*	cospectrum estimate (real part of cross-spectrum) of X and Y: $C_k^{xy} = \sum_{j=-p}^{p} W_j \mathrm{real}(\mathrm{J_{k+j}^{xy}})$ (except across endpoints)
QS_*nn_mm*	quadrature spectrum estimate (imaginary part of cross-spectrum) of X and Y: $Q_k^{xy} = \sum_{j=-p}^{p} W_j \mathrm{imag}(\mathrm{J_{k+j}^{xy}})$ (except across endpoints)
A_*nn_mm*	amplitude (modulus) of cross-spectrum of X and Y: $A_k^{xy} = \sqrt{(C_k^{xy})^2 + (Q_k^{xy})^2}$
K_*nn_mm*	coherency squared of X and Y: $K_k^{xy} = (A_k^{xy})^2/(F_k^x F_k^y)$
PH_*nn_mm*	phase spectrum in radians of X and Y: $\Phi_k^{xy} = \arctan(Q_k^{xy}/C_k^{xy})$

Printed Output

By default PROC SPECTRA produced no printed output.

When the WHITETEST option is specified, the SPECTRA procedure prints the following statistics for each variable in the VAR statement:

1. the name of the variable
2. M-1, the number of two-degree-of-freedom periodogram ordinates used in the tests
3. MAX(P(*)), the maximum periodogram ordinate
4. SUM(P(*)), the sum of the periodogram ordinates
5. Fisher's Kappa statistic
6. Bartlett's Kolmogorov-Smirnov test statistic

See "White Noise Test" earlier in this chapter for details.

ODS Table Names

PROC SPECTRA assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in the following table. For more information on ODS, see Chapter 6, "Using the Output Delivery System."

Table 17.2. ODS Tables Produced in PROC SPECTRA

ODS Table Name	Description	Option
WhiteNoiseTest	White Noise Test	WHITETEST
Kappa	Fishers Kappa	WHITETEST
Bartlett	Bartletts Kolmogorov-Smirnov Statistic	WHITETEST

Examples

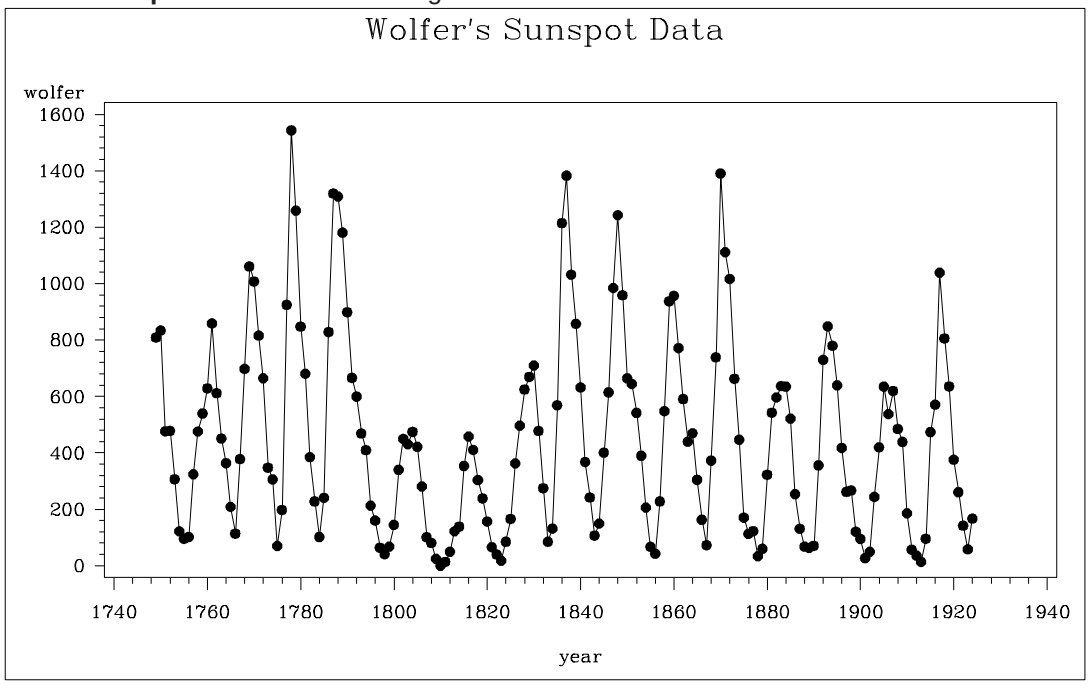
Example 17.1. Spectral Analysis of Sunspot Activity

This example analyzes Wolfer's sunspot data (Anderson 1971). The following statements read and plot the data.

```
title "Wolfer's Sunspot Data";
data sunspot;
   input year wolfer @@;
   datalines;
1749  809 1750  834 1751  477 1752  478 1753  307 1754  122 1755   96
1756  102 1757  324 1758  476 1759  540 1760  629 1761  859 1762  612
1763  451 1764  364 1765  209 1766  114 1767  378 1768  698 1769 1061
1770 1008 1771  816 1772  665 1773  348 1774  306 1775   70 1776  198
1777  925 1778 1544 1779 1259 1780  848 1781  681 1782  385 1783  228
1784  102 1785  241 1786  829 1787 1320 1788 1309 1789 1181 1790  899
1791  666 1792  600 1793  469 1794  410 1795  213 1796  160 1797   64
1798   41 1799   68 1800  145 1801  340 1802  450 1803  431 1804  475
1805  422 1806  281 1807  101 1808   81 1809   25 1810    0 1811   14
1812   50 1813  122 1814  139 1815  354 1816  458 1817  411 1818  304
1819  239 1820  157 1821   66 1822   40 1823   18 1824   85 1825  166
1826  363 1827  497 1828  625 1829  670 1830  710 1831  478 1832  275
1833   85 1834  132 1835  569 1836 1215 1837 1383 1838 1032 1839  858
1840  632 1841  368 1842  242 1843  107 1844  150 1845  401 1846  615
1847  985 1848 1243 1849  959 1850  665 1851  645 1852  542 1853  390
1854  206 1855   67 1856   43 1857  228 1858  548 1859  938 1860  957
1861  772 1862  591 1863  440 1864  470 1865  305 1866  163 1867   73
1868  373 1869  739 1870 1391 1871 1112 1872 1017 1873  663 1874  447
1875  171 1876  113 1877  123 1878   34 1879   60 1880  323 1881  543
1882  597 1883  637 1884  635 1885  522 1886  254 1887  131 1888   68
1889   63 1890   71 1891  356 1892  730 1893  849 1894  780 1895  640
1896  418 1897  262 1898  267 1899  121 1900   95 1901   27 1902   50
1903  244 1904  420 1905  635 1906  538 1907  620 1908  485 1909  439
1910  186 1911   57 1912   36 1913   14 1914   96 1915  474 1916  571
1917 1039 1918  806 1919  636 1920  376 1921  261 1922  142 1923   58
1924  167
;

symbol1 i=splines v=dot;
proc gplot data=sunspot;
   plot wolfer*year;
run;
```

The plot of the sunspot series is shown in Output 17.1.1.

Output 17.1.1. Plot of Original Data



The spectral analysis of the sunspot series is performed by the following statements:

```
proc spectra data=sunspot out=b p s adjmean whitetest;
   var wolfer;
   weights 1 2 3 4 3 2 1;
run;

proc print data=b(obs=12);
run;
```

The PROC SPECTRA statement specifies the P and S options to write the periodogram and spectral density estimates to the OUT= data set B. The WEIGHTS statement specifies a triangular spectral window for smoothing the periodogram to produce the spectral density estimate. The ADJMEAN option zeros the frequency 0 value and avoids the need to exclude that observation from the plots. The WHITETEST option prints tests for white noise.

The Fisher's Kappa test statistic of 16.070 is larger than the 5% critical value of 7.2, so the null hypothesis that the sunspot series is white noise is rejected.

The Bartlett's Kolmogorov-Smirnov statistic of 0.6501 is greater than

$$a\sqrt{1/(m-1)} = 1.36\sqrt{1/87} = 0.1458$$

so reject the null hypothesis that the spectrum represents white noise.

The printed output produced by PROC SPECTRA is shown in Output 17.1.2. The output data set B created by PROC SPECTRA is shown in part in Output 17.1.3.

Output 17.1.2. White Noise Test Results

```
                    Wolfer's Sunspot Data

                      SPECTRA Procedure

          Test for White Noise for Variable wolfer

                    M-1                87
                    Max(P(*))     4062267
                    Sum(P(*))    21156512


          Fisher's Kappa: (M-1)*Max(P(*))/Sum(P(*))

                      Kappa    16.70489


             Bartlett's Kolmogorov-Smirnov Statistic:
          Maximum absolute difference of the standardized
          partial sums of the periodogram and the CDF of a
                   uniform(0,1) random variable.

         Test Statistic                          0.650055
```

Output 17.1.3. First 12 Observations of the OUT= Data Set

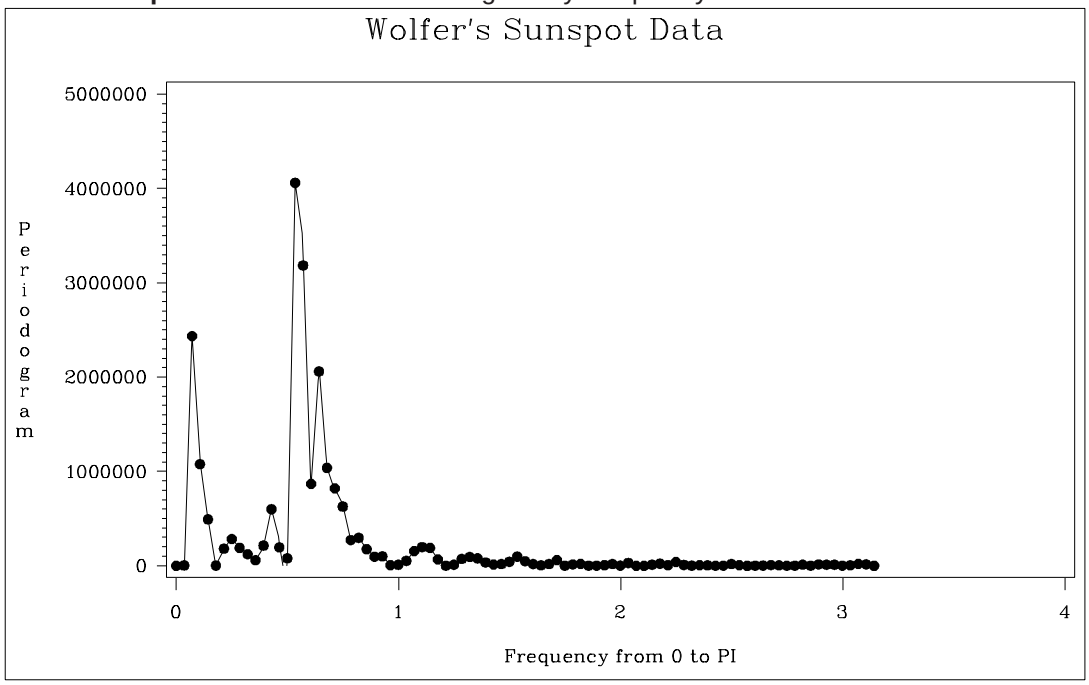
```
                    Wolfer's Sunspot Data

       Obs      FREQ      PERIOD          P_01      S_01

         1    0.00000         .             0.00    59327.52
         2    0.03570     176.000        3178.15    61757.98
         3    0.07140      88.000     2435433.22    69528.68
         4    0.10710      58.667     1077495.76    66087.57
         5    0.14280      44.000      491850.36    53352.02
         6    0.17850      35.200        2581.12    36678.14
         7    0.21420      29.333      181163.15    20604.52
         8    0.24990      25.143      283057.60    15132.81
         9    0.28560      22.000      188672.97    13265.89
        10    0.32130      19.556      122673.94    14953.32
        11    0.35700      17.600       58532.93    16402.84
        12    0.39270      16.000      213405.16    18562.13
```

The following statements plot the periodogram and spectral density estimate:

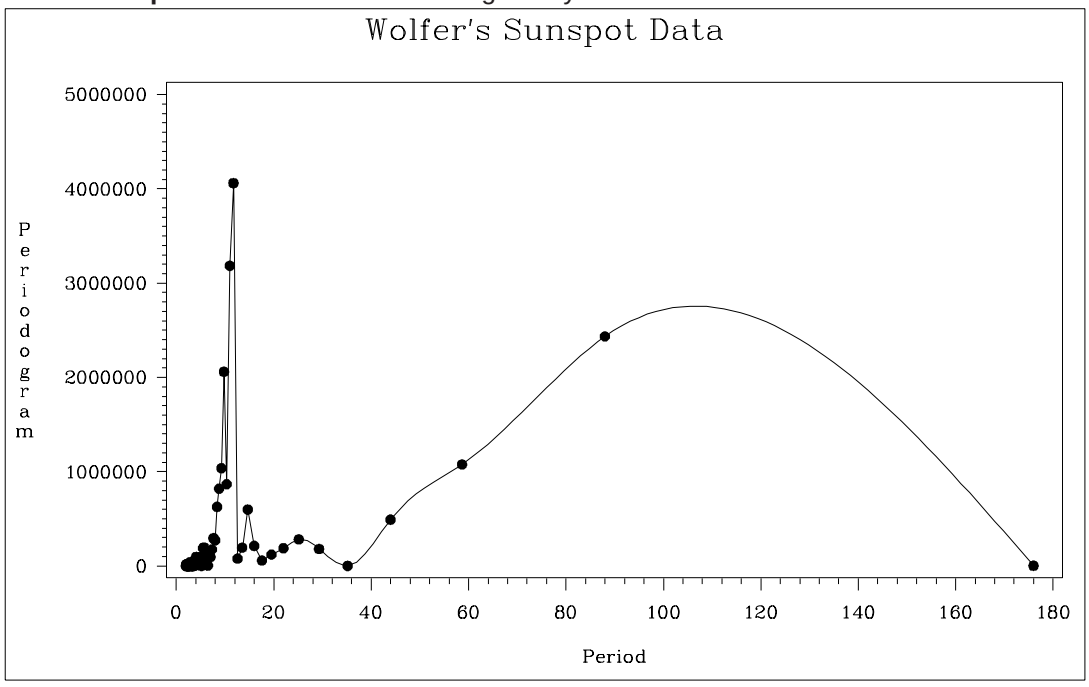
```
proc gplot data=b;
   plot p_01 * freq;
   plot p_01 * period;
   plot s_01 * freq;
   plot s_01 * period;
run;
```

The periodogram is plotted against frequency in Output 17.1.4 and plotted against period in Output 17.1.5. The spectral density estimate is plotted against frequency in Output 17.1.6 and plotted against period in Output 17.1.7.

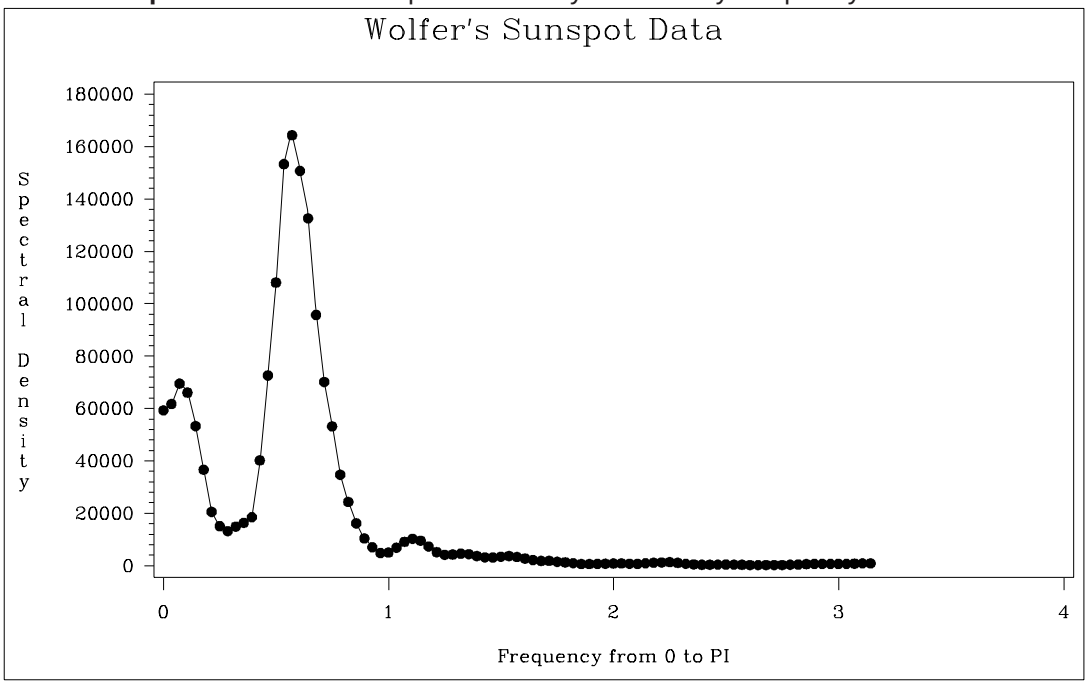
Output 17.1.4. Plot of Periodogram by Frequency



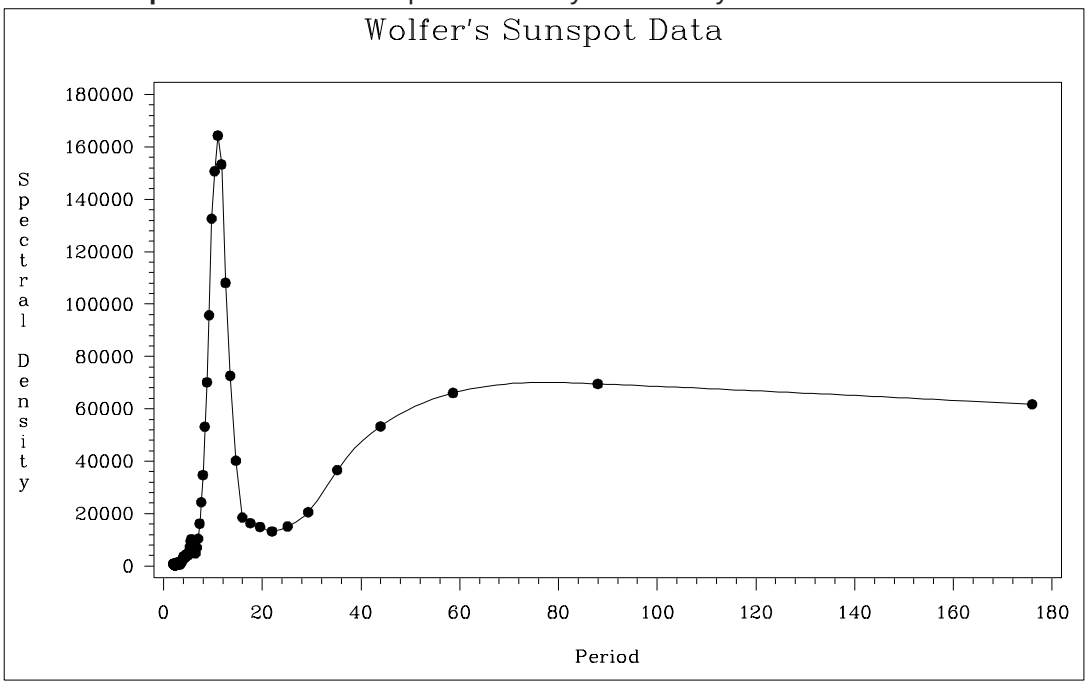
Output 17.1.5. Plot of Periodogram by Period



Output 17.1.6. Plot of Spectral Density Estimate by Frequency



Output 17.1.7. Plot of Spectral Density Estimate by Period



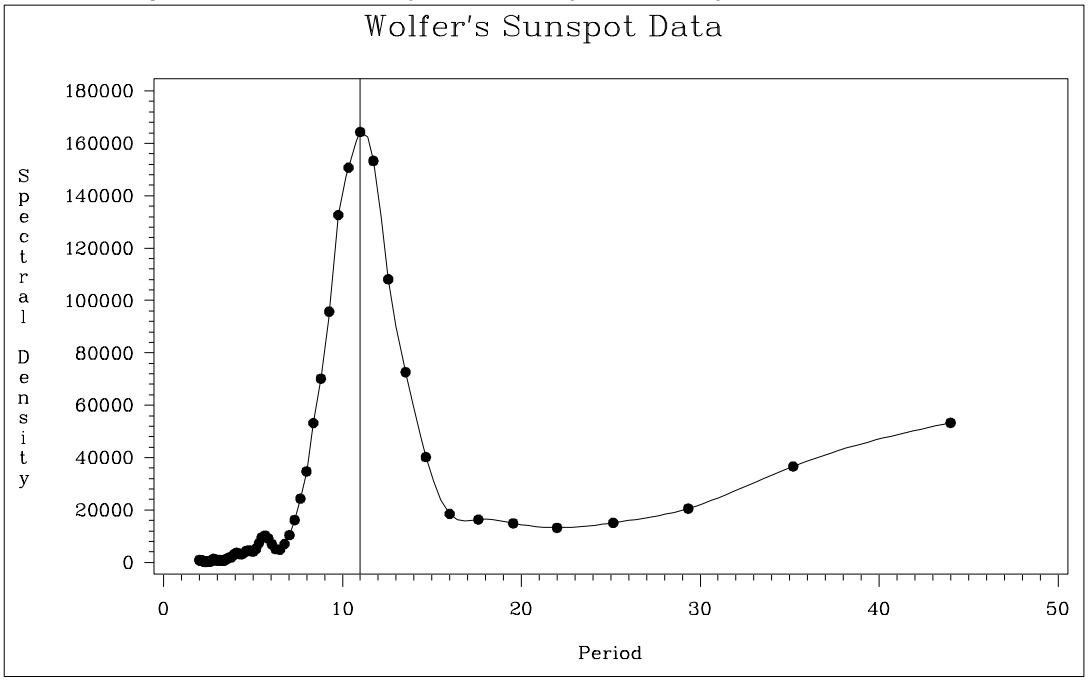
Since PERIOD is the reciprocal of frequency, the plot axis for PERIOD is stretched for low frequencies and compressed at high frequencies. One way to correct for this is to use a WHERE statement to restrict the plots and exclude the low frequency components. The following statements plot the spectral density for periods less than 50.

```
proc gplot data=b;
   where period < 50;
   plot s_01 * period / href=11;
run;
```

The spectral analysis of the sunspot series confirms a strong 11-year cycle of sunspot activity. The plot makes this clear by drawing a reference line at the 11 year period, which highlights the position of the main peak in the spectral density.

Output 17.1.8 shows the plot. Contrast Output 17.1.8 with Output 17.1.7.

Output 17.1.8. Plot of Spectral Density Estimate by Period to 50 Years



Example 17.2. Cross-Spectral Analysis

This example shows cross-spectral analysis for two variables X and Y using simulated data. X is generated by an AR(1) process; Y is generated as white noise plus an input from X lagged 2 periods. All output options are specified on the PROC SPECTRA statement. PROC CONTENTS shows the contents of the OUT= data set.

```
data a;
   xl = 0; xll = 0;
   do i = - 10 to 100;
      x = .4 * xl  + rannor(123);
      y = .5 * xll + rannor(123);
      if i > 0 then output;
      xll = xl; xl = x;
      end;
run;
```

```
proc spectra data=a out=b cross coef a k p ph s;
   var x y;
   weights 1 1.5 2 4 8 9 8 4 2 1.5 1;
run;

proc contents data=b position;
run;
```

The PROC CONTENTS report for the output data set B is shown in Output 17.2.1.

Output 17.2.1. Contents of PROC SPECTRA OUT= Data Set

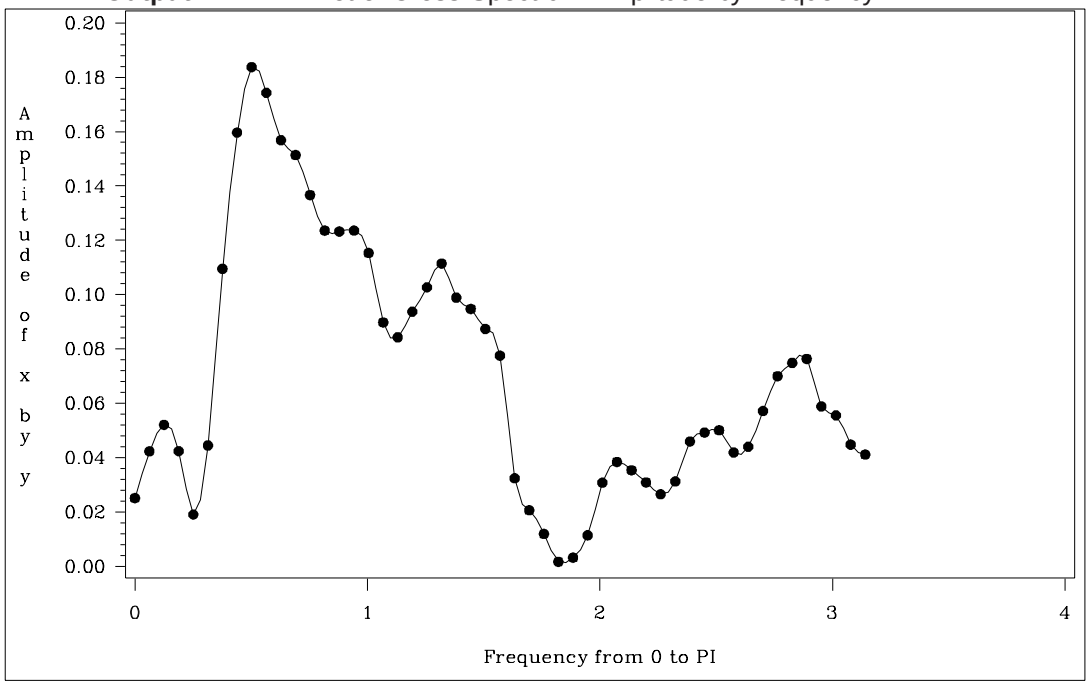
```
                          The CONTENTS Procedure

Data Set Name: WORK.B                                 Observations:         51
Member Type:   DATA                                   Variables:            17
Engine:        V8                                     Indexes:              0
Created:       12:39 Wednesday, April 28, 1999        Observation Length:   136
Last Modified: 12:39 Wednesday, April 28, 1999        Deleted Observations: 0
Protection:                                           Compressed:           NO
Data Set Type: DATA                                   Sorted:               NO
Label:         Spectral Density Estimates


                  -----Variables Ordered by Position-----

 #    Variable    Type    Len    Pos    Label
-----------------------------------------------------------------------
 1    FREQ        Num       8      0    Frequency from 0 to PI
 2    PERIOD      Num       8      8    Period
 3    COS_01      Num       8     16    Cosine Transform of x
 4    SIN_01      Num       8     24    Sine Transform of x
 5    COS_02      Num       8     32    Cosine Transform of y
 6    SIN_02      Num       8     40    Sine Transform of y
 7    P_01        Num       8     48    Periodogram of x
 8    P_02        Num       8     56    Periodogram of y
 9    S_01        Num       8     64    Spectral Density of x
10    S_02        Num       8     72    Spectral Density of y
11    RP_01_02    Num       8     80    Real Periodogram of x by y
12    IP_01_02    Num       8     88    Imag Periodogram of x by y
13    CS_01_02    Num       8     96    Cospectra of x by y
14    QS_01_02    Num       8    104    Quadrature of x by y
15    K_01_02     Num       8    112    Coherency**2 of x by y
16    A_01_02     Num       8    120    Amplitude of x by y
17    PH_01_02    Num       8    128    Phase of x by y
```

The following statements plot the amplitude of the cross-spectrum estimate against frequency and against period for periods less than 25.

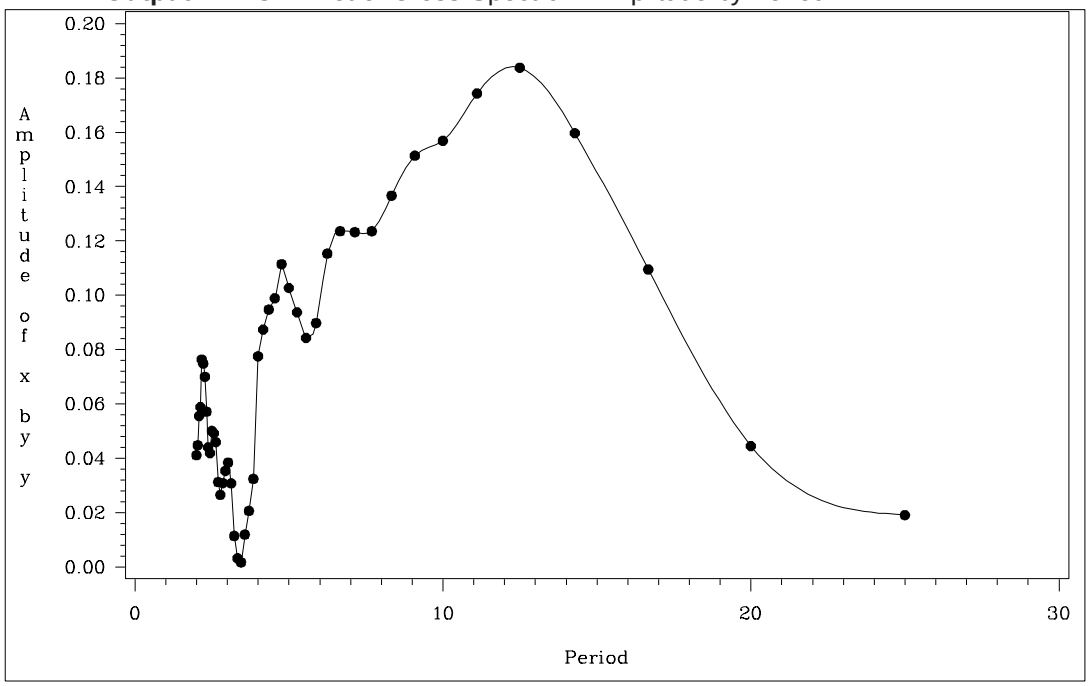
```
symbol1 i=splines v=dot;
proc gplot data=b;
   plot a_01_02 * freq;
run;

proc gplot data=b;
   plot a_01_02 * period;
   where period < 25;
run;
```

The plot of the amplitude of the cross-spectrum estimate against frequency is shown in Output 17.2.2. The plot of the cross-spectrum amplitude against period for periods less than 25 observations is shown in Output 17.2.3.

Output 17.2.2. Plot of Cross-Spectrum Amplitude by Frequency



Output 17.2.3. Plot of Cross-Spectrum Amplitude by Period



References

Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, New York: John Wiley & Sons, Inc.

Andrews, D.W.K. (1991), "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59 (3), 817-858.

Bartlett, M.S. (1966), *An Introduction to Stochastic Processes*, Second Edition, Cambridge: Cambridge University Press.

Brillinger, D.R. (1975), *Time Series: Data Analysis and Theory*, New York: Holt, Rinehart and Winston, Inc.

Davis, H.T. (1941), *The Analysis of Economic Time Series*, Bloomington, IN: Principia Press.

Durbin, J. (1967), "Tests of Serial Independence Based on the Cumulated Periodogram," *Bulletin of Int. Stat. Inst.*, 42, 1039–1049.

Fuller, W.A. (1976), *Introduction to Statistical Time Series*, New York: John Wiley & Sons, Inc.

Gentleman, W.M. and Sande, G. (1966), "Fast Fourier transforms–for fun and profit," *AFIPS Proceedings of the Fall Joint Computer Conference*, 19, 563–578.

Jenkins, G.M. and Watts, D.G. (1968), *Spectral Analysis and Its Applications*, San Francisco: Holden-Day.

Miller, L. H. (1956), "Tables of Percentage Points of Kolmogorov Statistics," *Journal of American Statistic Association*, 51, 111.

Monro, D.M. and Branch, J.L. (1976), "Algorithm AS 117. The chirp discrete Fourier transform of general length," *Applied Statistics*, 26, 351–361.

Nussbaumer, H.J. (1982), *Fast Fourier Transform and Convolution Algorithms*, Second Edition, New York: Springer-Verlag.

Owen, D. B. (1962), *Handbook of Statistical Tables*, Addison Wesley.

Parzen, E. (1957), "On Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics*, 28, 329-348.

Priestly, M.B. (1981), *Spectral Analysis and Time Series*, New York: Academic Press, Inc.

Singleton, R.C. (1969), "An Algorithm for Computing the Mixed Radix Fast Fourier Transform," *I.E.E.E. Transactions of Audio and Electroacoustics*, AU-17, 93–103.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/ETS User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999. 1546 pp.

SAS/ETS User's Guide, Version 8

ISBN 1–58025–489–6

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999