Tutorial

Using the SAS System to Perform Univariate and Cross-Spectral Analysis

John C. Brocklebank
SAS Institute Inc.
Cary, N.C.

David A. Dickey
Associate Professor of Statistics
North Carolina State University
Raleigh N.C.

ABSTRACT

Time series analysis is widely used in engineering, the physical sciences, economics, and statistics. Spectral analysis, or the "frequency domain" approach to time series analysis, is popular in engineering applications. Firmly intertwined with the spectral approach is the "correlations" or "time domain" approach, often used for statistical modeling.

Many find the concepts of powers, frequencies, and periodicities in spectral analysis difficult to grasp. Yet with applications covering many fields and a wide range of problems, spectral analysis proves to be a fascinating field of study.

OVERVIEW

This tutorial focuses on the "spectral" approach to time series analysis using the SAS System. The many relationships between the spectral approach and the "time" domain approach are discussed. The tutorial covers the following topics:

- Definition of spectrum
- Spectra of ARMA processes, including white noise
- Definition of periodogram and distribution of periodogram ordinates
- Obtaining plots of the periodogram and smoothed periodogram from the SPECTRA procedure
- Considerations for choosing appropriate weights for smoothing
- Testing for hidden periodicities
- Multiple series and cross-spectral quantities

I. PERIODIC DATA

* Spectral analysis was developed to detect sinusoidal components in time series models.

Consider a model

\[ y_t = \mu + \alpha \sin(\omega t) + \epsilon_t \]

where \( \epsilon_t \) is a sequence of uncorrelated \((0, \sigma^2)\) variables and the amplitude \( \alpha \) is small relative to the variance of \( \epsilon_t \).

* One can find \( \alpha \) and \( \mu \) such that

\[ \alpha = \sin(\omega \Delta t) \]

* Given the frequency \( \omega \), perform a regression of \( y_t \) on \( \sin(\omega t) \), \( \cos(\omega t) \) and the overall F test considers \( \alpha = 0 \) (no component at frequency \( \omega \)).

Example

- Given water usage data in a household measured over an eight week period,
- Goal is to determine whether or not there is a weekly cycle in water usage.
- Plot the original data.

![Household Water Usage](image)

Example continued

- Use the GLM procedure to determine if there is a significant cycle at \( \omega = 2\pi/7 \), which corresponds to a weekly cycle (2\pi radians over 7 days).
- Create SAS variables in a DATA step

\[
\text{SIN} = \sin(2 * 3.14159 * \text{N}/7);
\]

\[
\text{COS} = \cos(2 * 3.14159 * \text{N}/7);
\]

- Use PROC GLM to test \( \alpha = 0 \) using the F test

\[
\text{F} = \frac{\text{MSBETWEEN}}{\text{MSRESIDUAL}}
\]

- The regression shows a significant cycle at 2\pi/7, not visible to the naked eye since the amplitude of this sinusoid is small compared to the variance of the noise term.
Example continued

Since the frequency \( \omega \) may not be known you can use PROC SPECTRA to search for hidden periodicities by running regressions on a sequence of \( \omega \) values.

The sums of squares are equivalent to those obtained by regressing the variable of interest on sine and cosine variables at frequencies 
\[ \omega_j = \frac{2\pi j}{n}, \quad j = 0, 1, \ldots, n/2. \]

The values of \( \omega_j \) are called the Fourier frequencies (efficient computationally).

Periogram

We examine the periodogram ordinates using

- PROC SPECTRA for our water heater data.
- We use the ADJMEAN option to remove the effect of mean ordinate which would destroy the resolution of our plot.
III. FORM OF THE SPECTRUM

- The spectrum is a theoretical quantity that the normalized periodogram ordinates are estimating for the model

\[ Y_t = \sum_{i=1}^{k} (A_i \sin(w_i t) + B_i \cos(w_i t)) + \epsilon_t \]

with \( A_i, B_i \) independent \( N(0, a_i^2) \) and \( \epsilon_t \) uncorrelated sequence.

- One expects a flat spectrum except at frequencies \( w_i \), \( i = 1, \ldots, k \).

- Interpretation of the spectrum at frequency \( w \) is the amount of overall variation in \( Y \) associated with frequency \( w \).

- For ARIMA models, the spectrum can be calculated as the Fourier transform of the autocovariance function.

- For a real series, the spectrum is defined by the continuous function

\[ f(\omega) = \frac{1}{\pi} \sum_{n=-\infty}^{\infty} \gamma(n \omega) \cos(n \omega) \]

where \( \omega \) is the frequency in radians, and is in the range \((-\pi, \pi)\).

- Note for a white noise (i.e., \( \gamma(T) = 0 \) for \( T \) different from 0), \( f(\omega) = \omega^2/(2\pi) \) for all \( \omega \) and hence the spectrum is flat.

Spectrum for Different Series

- Stationary series with positive autocorrelation tend to have smooth spectra with high ordinates or power at low frequencies and low ordinates at high frequencies.

- Stationary series with negative autocorrelation are smooth display power at the high frequencies and low ordinates at the low frequencies.

- The spectra of stationary series with negative autocorrelation are smooth display power at the high frequencies and low ordinates at the low frequencies.
IV. SMOOTHING THE PERIODGRAM

- The spectrum is a smooth function for many time series but the periodogram is highly volatile.

- We can smooth the periodogram by using a weighted moving average.

- This is justified by assuming the periodogram ordinates at neighboring frequencies are estimating the same spectral ordinate.

- Weight functions concentrated on a few frequencies tend to improve the resolution of the peaks in the estimated spectrum and are described as having narrow band widths (don’t smooth over peaks very much).

- Weight functions giving substantial weight to several frequencies tend to smooth over the peaks and are said to have wide band widths. (Some resolution is lost since sharp peaks now appear as smoother “hills” and two very close peaks can be smoothed over into a single peak.)

- Weight functions of different band widths are usually used and several plots inspected.

- Sometimes tapered weight sequences, like 1 2 3 2 1, are used instead of straight averages, like 1 1 1 1.1.

- The set of weights is called a spectral window.

- The format of the weights statement in PROC SPECTRA is:

  WEIGHTS list of weights;

Smoothing the Periodogram

- For a nonconstant spectrum, the weighted averages usually have some bias.

- The larger band width weight sequences tend to have more bias (but less variance).

For our water heater example, we execute these SAS statements:

```sas
PROC SPECTRA S;
VAR Y;
WEIGHTS 1 3 5 2 1;
PROC GPLOT;
PLOT S_OI,"REQ;
```

Smoothing the Periodogram

- For several series, we compute the periodogram and the smoothed spectral estimates.

- The spectral density plots are overlaid on the true spectra.

\[ Y(t) = 0.7Y(t-1) + E(t) \]
V. TESTS FOR WHITE NOISE

- Symbol \( e_t \) represents uncorrelated random variables with mean 0 and variance \( \sigma^2 \).
- Typically call \( e_t \) "white noise".
- ARIMA models are defined in terms of white noise.
- Given \( e_1, e_2, \ldots, e_n \), best prediction of \( e_{n+1} \) is zero (uncorrelated property).
- Good models decompose observations into predictable parts in addition to white noise.
- White noise is a basic building block for ARIMA Models.

White Noise Tests

- Fuller (1976) describes the tests for the null hypothesis that a series is white noise.
- Fisher-Kappa Test is the ratio of the largest periodogram ordinate to the average of the ordinates.
- This statistic is compared to the table in Fuller.
- The test is designed to detect one sinusoidal component buried in white noise.
- For other departures from the white-noise hypothesis, some prefer the Bartlett Kolmogorov-Smirnov test that accumulates departures from the white-noise hypothesis over all frequencies.
- This test is based on the cumulative (summed) periodogram, and the ratios calculated are input for a standard Kolmogorov-Smirnov test.

Example Continued

As an example using our water usage series, we can run the following SAS statements.

```
PROC SPECTRA P=WHITETEST;
VAR Y;
```

The 5% critical value for the Fisher-Kappa Test is 3.915 and the 5% critical value for the Kolmogorov-Smirnov test is 2.5.

VI. CROSS-SPECTRAL ANALYSIS

- Interpretation of cross-spectral quantities is intimately related to the transfer function model in which an output time series, \( Y_t \), is related to an input series, \( X_t \), through the equation
  \[
  Y_t = \sum_{j=0}^{\infty} \psi_j X_{t-j} + e_t
  \]
  and \( e_t \) is a white noise series independent of the input \( X_t \).
- For the moment assume \( e_t = 0 \).
Cross Spectral Analysis

- For example, let $Y_t$ be the deviation of room temperature from 68 degrees and $X_t$ be the deviation of the logarithm of furnace temperature from five.
- Suppose our transfer function is

$$Y_t = A Y_{t-1} - X_t,$$

then

$$Y_t = 1.8 Y_{t-1} - 1.8^2 X_t,$$

is a weighted sum of current and previous inputs.
- Cross-spectral quantities tell you what happens to sinusoidal inputs.
- Suppose $X_t$ (or a component thereof) is the sinusoid

$$X_t = \sin (w t),$$

where $w = 2\pi/12$.

Cross-Spectral Analysis

- It is known that $Y_t$ satisfying

$$Y_t = 0.8 Y_{t-1} + \sin (w t);$$

must be of the form $Y_t = A \sin (w t - \theta)$.
- Solving we obtain $A = 1.3022$ and $\theta = 1.98$.
- The interpretation is that an input sinusoid produces an output with amplitude $1.98$ times that of the input; it has the same frequency and phase shift $\text{Arctan}(1.3022) = 22.5^\circ$.
- The output for any noiseless linear transfer function is a sinusoid of frequency $w$ when the input $X$ is such a sinusoid.
- Only the amplitude and phase are changed.

Cross-Spectral Analysis

- Cross-spectral analysis estimates the gain and phase at all Fourier frequencies using arbitrary input and associated output.
- Intermediate here is the computation of variety of other useful quantities.
- The cospectrum $c(w)$ and the quadrature spectrum $q(w)$ represent the real and imaginary part of the cross spectrum $f_{xy}(w)$.
- The cross-amplitude spectrum is defined as

$$A_{xy}(w) = |f_{xy}(w)| = c^2(w) + q^2(w)^{1/2} f_{xy}(w).$$
- The gain is defined as

$$A_{xy}(w)/f_{xx}(w).$$
- The phase spectrum $\phi_{xy}(w)$ is defined as

$$\phi_{xy}(w) = \text{Arctan}(q(w)/c(w)).$$

Cross-Spectral Analysis

- The cross spectrum is thus expressed as

$$f_{xy}(w) = A_{xy}(w) \exp(i \phi_{xy}(w)).$$
- Introducing an error series into the model, one obtains

$$Y_t = 1.8 Y_{t-1} + \sin (w t) + N_t.$$

- Similar to a correlation coefficient the squared coherency measures the strength of the relationship between $XY$ and is represented by

$$K_{xy}^2(w) = f_{xy}(w)/f_{xx}(w) f_{yy}(w).$$
- The spectrum $f_{xy}(w)$ of $X_t$ satisfies

$$f_{xy}(w) = f_{yx}(w)(1-K_{xy}^2(w)).$$

Example

The plots below examine the true and estimated cross-spectral estimates for 512 observations $Y_t$ derived from the following process:

$$Y_t = 0.8 Y_{t-1} + X_t,$$

where $X_t$ is an autoregressive process $X_t = 0.5 X_{t-1} + \varepsilon_t$ with variance 1.333 and $Y_t$ is white noise with variance one.
- The $X_t$ represents fluctuations in the furnace temperature and $Y_t$ fluctuations in room temperature.
- The model assumes no feedback (there is no thermostat in the room).

Interpretation of Phase Spectra

- The phase spectrum shows that the furnace and room temperature fluctuations, for very long period fluctuations (near 0) and very short periods (near 1), are nearly in phase.
- The phase spectrum starts at 0 and then decreases, indicating that $X$ (the furnace temperature) tends to hit its peak slightly before the room temperature.

Interpretation of the Squared Coherency

- The squared coherency is near one at the low frequencies, indicating a strong correlation between room and furnace temperatures at the low frequencies.
The squared coherence becomes smaller at higher frequencies and the phase spectrum is quite variable at the high frequencies as a result of the low correlation between furnace and room temperature.

Cross Amplitude Spectrum
- The gain behaves like the cross-amplitude spectrum in this example.
- This shows that low frequency fluctuations in the furnace produce high amplitude fluctuations in the room temperature.
- High frequency fluctuations in the furnace produce low amplitude (low variance) fluctuations in the room temperature.

Cross Amplitude Y By X

VII. SUMMARY
- Spectral analysis is used to find hidden periodicities in data.
- Shapes of spectra also indicate the nature of autocorrelation in stationary models.
- A flat spectrum indicates white noise. Hence spectral analysis can be used as a check on residuals.
- Cross-spectral analysis relates inputs and outputs on a frequency-by-frequency basis.

SAS is the registered trademark of SAS Institute Inc., Cary, NC, USA

REFERENCES