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 power, 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

  $Y_t = \mu + \alpha \sin(\omega_t + \delta) + e_t$

  where

  e_t is a sequence of uncorrelated $(0, \sigma^2)$ variates and the amplitude α is small relative to the variance of e_t.

- One can find A and B such that

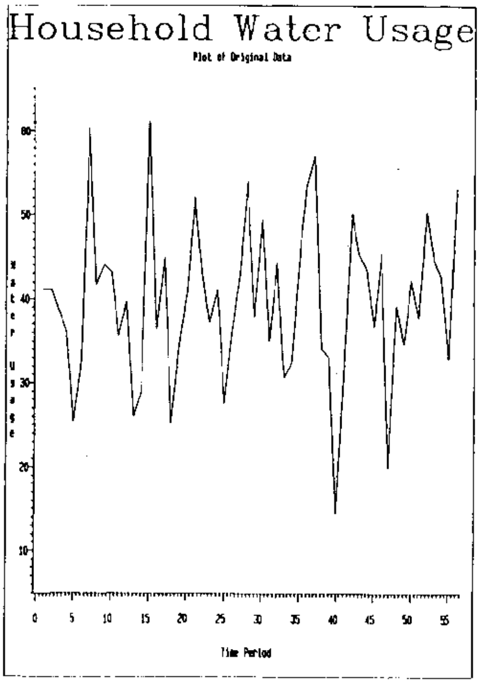
  $\alpha \sin(\omega_t + \delta) = A\sin(\omega_t) + B\cos(\omega_t)$

- Given the frequency ω, perform a regression of Y_t on 1, $\sin(\omega_t)$, $\cos(\omega_t)$ and the overall F test considers $H_0: \alpha = 0$ (no component at frequency ω).

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.

```
PROC GPLOT;
   PLOT Y*T;
```



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 (2Π radians over 7 days).
- Create SAS variables in a DATA step

```
SIN = SIN(2 * 3.14159 *  N /7);

COS = COS(2 * 3.14159 *  N /7);
```

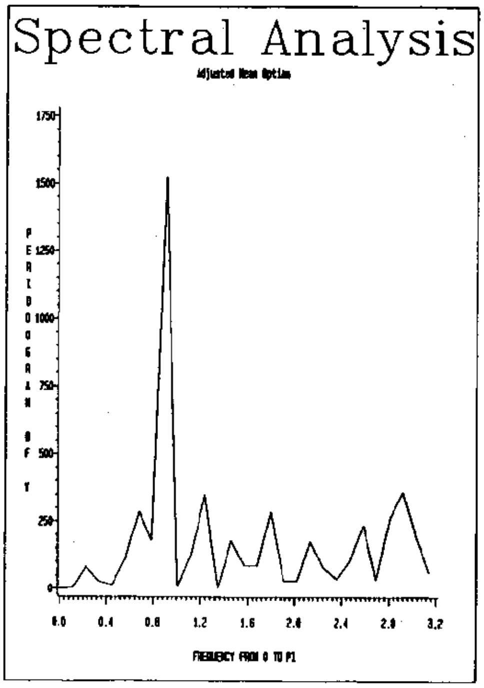
- Use PROC GLM to test $\alpha=0$ using the F test

```
PROC GLM;
   MODEL Y=SIN COS;
```

- 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

```
                 Regression on Sine and Cosine Terms

                  GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE              DF      SUM OF SQUARES            MEAN SQUARE

MODEL                2        1523.58881727           761.79440863

ERROR               53        3407.33829569            64.28940181

CORRECTED TOTAL     55        4930.92711296

MODEL F =        11.85                          PR > F = 0.0001

R-SQUARE          C.V.            ROOT MSE                 Y MEAN

0.308986       20.1274          8.01806721             39.83656237

SOURCE              DF          TYPE I SS    F VALUE     PR > F

SIN                  1      1207.35434403      18.78     0.0001
COS                  1       316.23447320       4.92     0.0309

SOURCE              DF        TYPE III SS    F VALUE     PR > F

SIN                  1      1207.35432764      18.78     0.0001
COS                  1       316.23447320       4.92     0.0309

                                T FOR H0:     PR > |T|    STD ERROR OF
PARAMETER         ESTIMATE    PARAMETER=0                   ESTIMATE

INTERCEPT      39.83656228          37.18       0.0001      1.07165930
SIN             6.56656667           4.33       0.0001      1.51527228
COS             3.39066782           2.22       0.0309      1.51527227
```



II. PERIODOGRAM

- Since the frequency ω may not be known you can use PROC SPECTRA to search for hidden periodicites by running regressions on a sequence of ω values.
- The sums of squares are equivalent to those obtained by regressing the variable of interest on sine and cosine variables at frequencies

  $\omega_j = 2\Pi j/n \quad j=0,1,\ldots,n/2.$
- The values of ω_j are called the Fourier frequencies (efficient computationally).

Periodogram

We examine the periodogram ordinates using

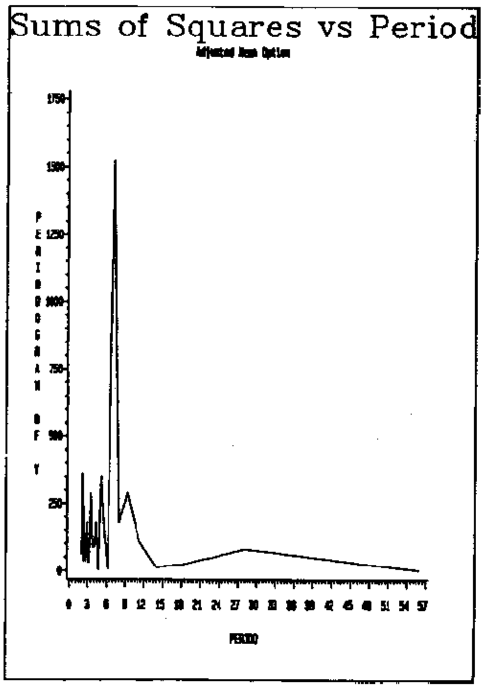
- PROC SPECTRA for our water heater data.
- We use the ADJMEAN option to remove the effect of mean ordinate which would destroy the resolution of our plot.

```
PROC SPECTRA P ADJMEAN;
   VAR Y;
PROC GPLOT;
   PLOT P_01*FREQ;
```

Plotting Sum of Squares Against Period

We can also plot the sums of squares against period using the SAS statement

```
PLOT P_01* PERIOD;
```



III. FORM OF THE SPECTRUM

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

  $Y_t = \Sigma_{i=1,k}(A_i \sin(\omega_i t) + B_i \cos(\omega_i t)) + e_t$

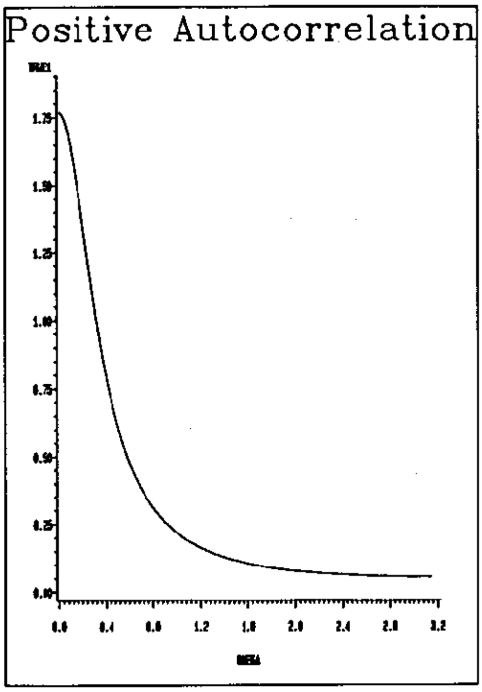
  with A_i, B_i independent $N(0,\sigma_i^2)$ and e_t uncorrelated sequence.
- One expects a flat spectrum except at frequencies ω_i, $i=1,\ldots,k$.
- Interpretation of the spectrum at frequency ω is the amount of overall variation in Y associated with frequency ω.
- For ARMA models, the spectrum can be calculated as the Fourier transform of the autocovariance function.
- For a real series, the spectrum is defined by the continuous function

  $f(\lambda)=(2\pi)^{-1}(\gamma(0)+2\Sigma_{T=1,\infty}\ \gamma(T)\cos(\lambda T))$

  where λ 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), $f(\lambda)=\sigma^2/(2\pi)$ for all λ and hence the spectrum is flat.

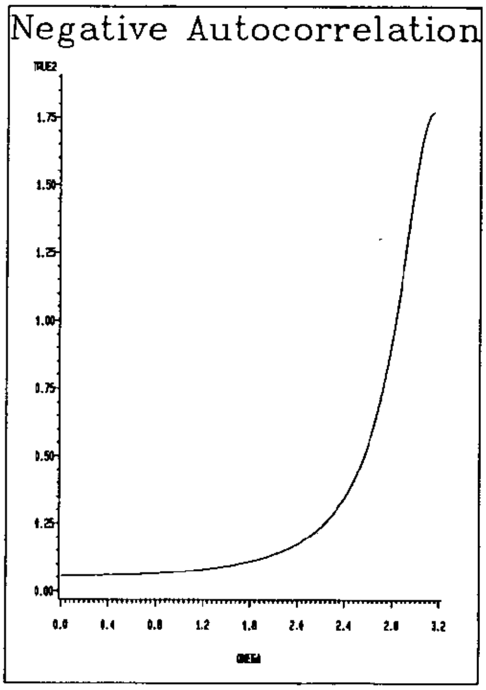
Spectrum for Different Series

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



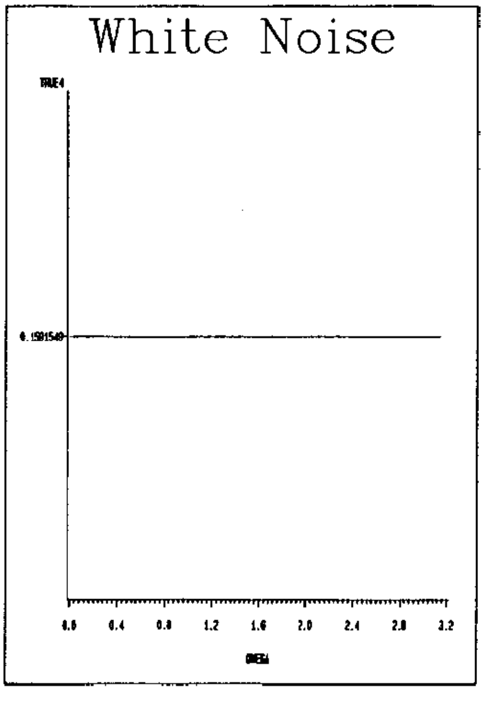
Spectrum for Different Series

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



Spectrum for Different Series

- A series with an exact sinusoidal component at ω_0 has a spike in the spectrum at frequency ω_0.
- A white noise series has flat spectra. This makes the spectrum a useful tool for diagnostic checking of model residuals.



IV. SMOOTHING THE PERIODOGRAM

- 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;
```

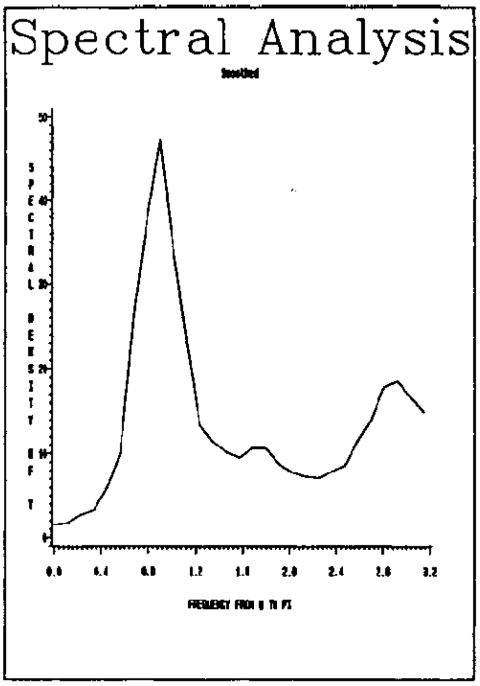
Smoothing the Periodogram

- 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

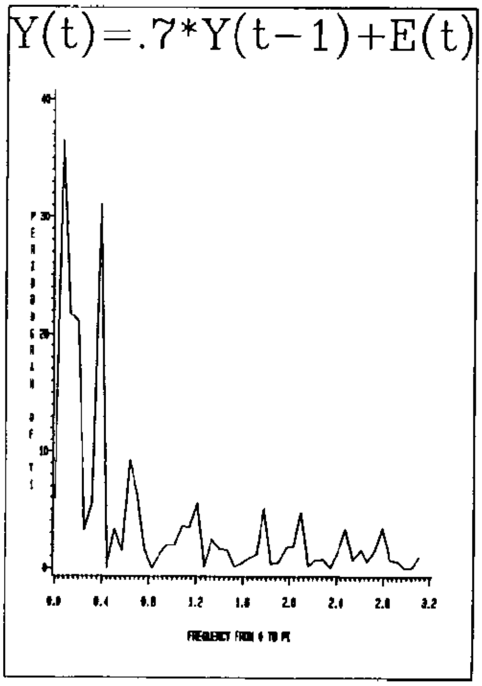
```
PROC SPECTRA S;
   VAR Y;
   WEIGHTS 1 2 3 2 1;
PROC GPLOT;
   PLOT S_01*FREQ;
```

Smoothing the Periodogram

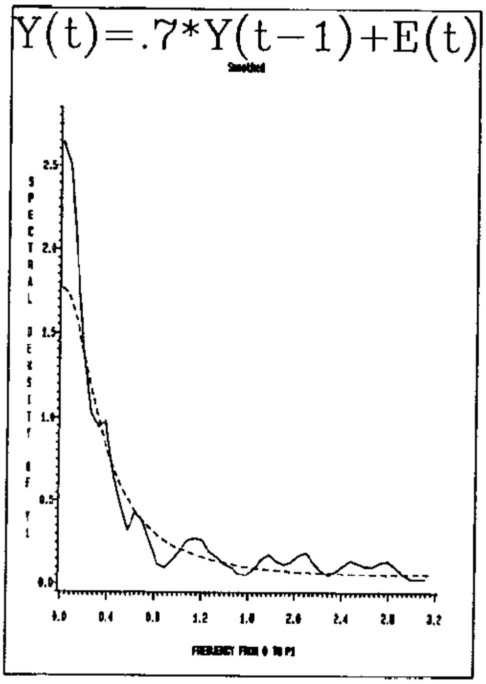


Smoothing the Periodogram

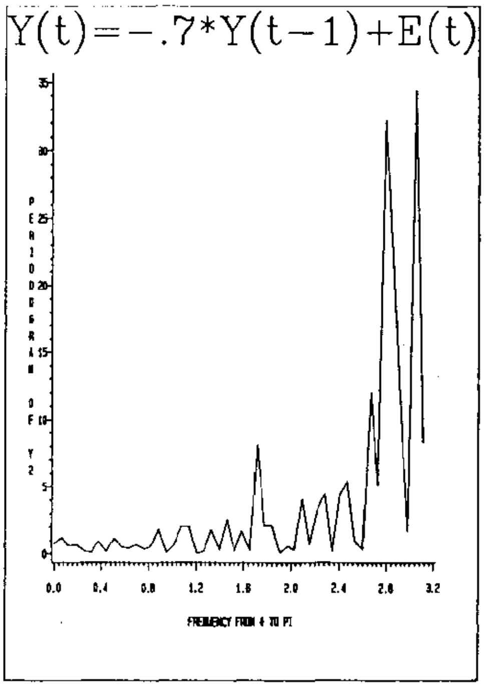
- For several series, we compute the periodogram and the smoothed spectral estimates.
- The spectral density plots are overlaid on the true spectra.



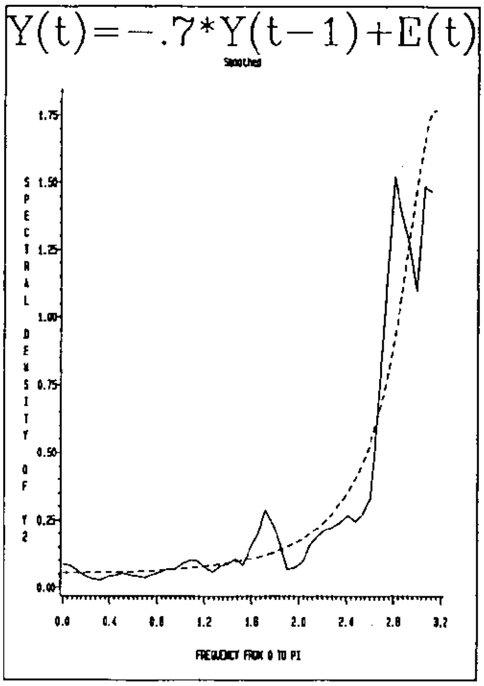
Smoothing the Periodogram



Smoothing the Periodogram



Smoothing the Periodogram



V. TESTS FOR WHITE NOISE

- Symbol e_t represents uncorrelated random variables with mean 0 and variance σ^2.
- Typically call e_t "white noise".
- ARIMA models are defined in terms of white noise.
- Given e_1, e_2, ..., e_n, best prediction of e_{n+j} 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 5.915 and the 5% critical value for the Kolmogorov-Smirnov test is .25.

```
                    Household Water Usage
                      White Noise Test

-----  TEST FOR WHITE NOISE FOR VARIABLE Y          -----

FISHER'S KAPPA: (M-1)*MAX(P(*))/SUM(P(*))

PARAMETERS:      M-1         =        27
                 MAX(P(*)) =  1523.589
                 SUM(P(*)) =  4900.426

TEST STATISTIC: KAPPA      =    8.3946

BARTLETT'S KOLMOGOROV-SMIRNOV STATISTIC:
         MAX ABS DIFFERENCE OF THE STANDARDIZED PARTIAL SUMS
         OF THE PERIODOGRAM AND THE CDF OF A UNIFORM(0,1).

TEST STATISTIC =    0.1835
```

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

  $Y_t = \Sigma_{j=-\infty,\infty} V_j X_{t-j} + n_t$

  and n_t is a white noise series independent of the input X_t.
- For the moment assume $n_t = 0$.

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

  $Y_t - .8\, Y_{t-1} = X_t$.

  then

  $Y_t = \Sigma_{j=0,\infty} (.8)^j X_{t-j}$

  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 sinuisoid

  $X_t = \sin(\omega_t)$

  where

  $\omega = 2\Pi/12$.

Cross-Spectral Analysis

- It is known that Y_t satisfying

  $Y_t - .8\, Y_{t-1} = \sin(\omega_t)$

  must be of the form $Y_t = A\sin(\omega_t - B)$.
- Solving we obtain tan (B) = 1.3022 and A = 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 Arctan (1.3022) = 52.5°.
- The output for any noiseless linear transfer function is a sinusoid of frequency ω 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 $c(\omega)$ and the quadrature spectrum $q(\omega)$ represent the real and imaginary part of the cross spectrum $f_{xy}(\omega)$.
- The cross-amplitude spectrum is defined as

  $A_{xy}(\omega) = |f_{xy}(\omega)| = (c^2(\omega) + q^2(\omega))^{.5} f_{xx}(\omega)$.
- The gain is defined as

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

  $\Psi_{xy}(\omega) = \text{Atan}(q(\omega)/c(\omega))$.

Cross-Spectral Analysis

- The cross spectrum is thus expressed as

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

  $Y_t = \Sigma_{j=-\infty,\infty} V_j X_{t-j} + \eta_t$.

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

  $K^2_{xy}(\omega) = |f_{xy}(\omega)|^2/(f_{xx}(\omega) f_{yy}(\omega))$.
- The spectrum $f_\eta(\omega)$ of η_t satisfies

  $f_\eta(\omega) = f_{yy}(\omega)(1 - K^2_{xy}(\omega))$.

Example

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

$V_t = .8\, V_{t-1} + X_t$

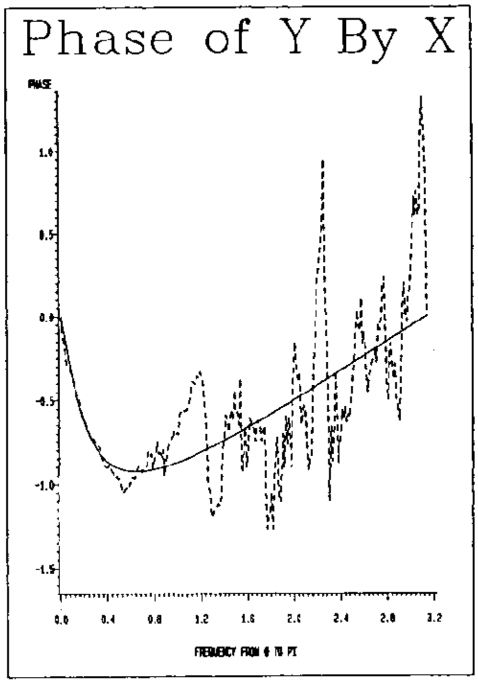
$Y_t = V_t + \eta_t$

where X_t is an autoregressive process $X_t = .5\, X_{t-1} + e_t$ with variance 1.3333 and η_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

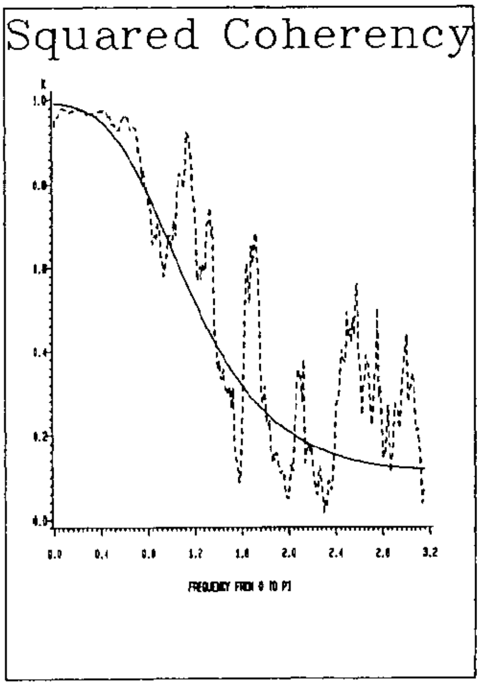
- The phase spectrum shows that the furnace and room temperature fluctuations, for very long period fluctuations (ω near 0) and very short periods (ω near Π), are nearly in phase
- The phase spectrum starts at 0 and then decreases, indicating that X (the furnace temperature) tends to hit its peak slightly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



Interpretation of the Squared Coherency

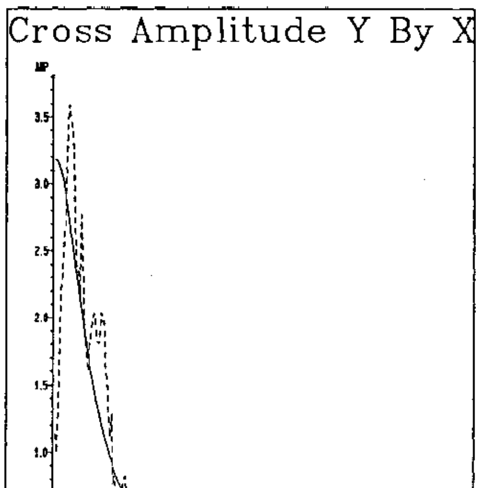
- The squared coherency is near one at the low frequencies, indicating a strong correlation between room and furnace temperatures at the low frequencies.

- The squared coherency becomes smaller at higher frequencies and the phase spectrum is quite variable at the high frequencies as a result of this 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 produces low amplitude (the variance) fluctuations in the room temperature.



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.

REFERENCES

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

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, New York: John Wiley & Sons.

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

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

Priestly, M.B. (1981), Spectra Analysis and Time Series, I, New York: Academic Press.

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.