

Case Studies in Time Series: Periodic Behavior and Related Issues

David A. Dickey
North Carolina State University, Dept. Of Statistics

INTRODUCTION

Many SAS users and most econometricians think of time series strictly in the time domain, that is, the series is a set of observations written Y_t , where t represents time $t=1,2, \dots$ and the task of time series modelers is to predict future values of Y from past observed Y values along with, possibly, past observed values of other predictor series. However, there are many time series, for example tidal movements, environmental variables, and even economic series that can be more conveniently modeled in the so-called frequency domain. That is, some time series can logically be thought of as having deterministic periodic components that can be extrapolated into the future to produce predictions. Without regard to forecasting, the analysis of periodic data, such as the noise waves emitted by a submarine, can be of interest in themselves. In fact forecasting is often of little or no interest in these cases. We might be able to associate a particular wave form with a particular class of submarine, for example. Analysis of light waves absorbed by or passed through a liquid can be of help in identifying the liquid's components. In economics one can think of long term and short term fluctuations by breaking the series into periodic components. An analysis in which a time series is broken down into periodic components goes by the names spectral analysis, periodogram analysis, or in more general mathematical terms harmonic analysis or Fourier analysis. This article is an introductory exploratory investigation of this method.

AN EXAMPLE

Let's start with an artificial example. The graph below represents the thickness of a coating applied to fabric as it passes through a series of rollers. The rollers have circumferences 5, 6, 8, 11, 16 and 20 inches. If one of the rollers is slightly off center, one expects this thickness to rise and fall as that roller completes each cycle. This is analogous to the vibration that would be caused in an automobile tire that is out of balance or the wiggling observed in an improperly balanced ceiling fan or washing machine. In our example we want to analyze these measurements to see if there is any periodic behavior, and if so, which of the rollers is involved. Each period identified will determine the circumference of exactly one of the rollers. Careful measurements of thickness, Y , are made every quarter of an inch. Measurements are in thousandths of an inch and the target thickness is 10. There are 512 observations. The data plotted in Figure 1.

```
proc gplot data=fabric; plot Y*location;  
  symbol1 v=none i=join c=blue;
```

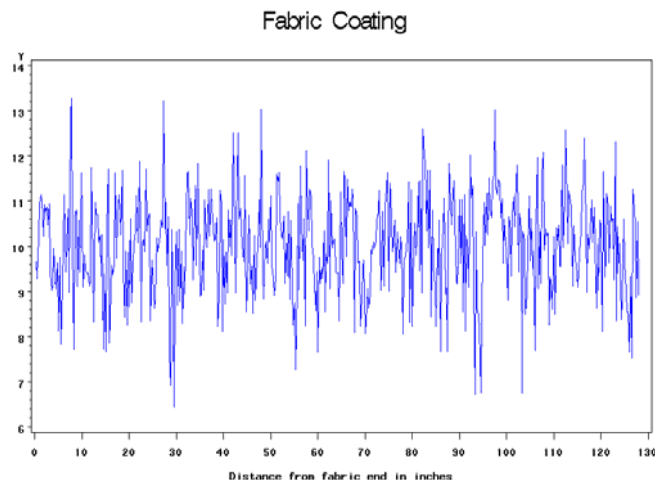


Figure 1: Fabric coating thickness every quarter inch.

No cycles appear to the naked eye. Next we describe and then compute the "periodogram," whose horizontal axis represents frequency in radians per observation. What is a radian? It is a measure for angles. Imagine running a string as long as a circle's radius along the circle's edge. Draw a line from each end of the string to the center of the circle. The angle between those lines is a radian.

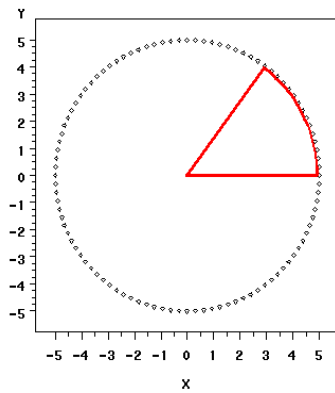


Figure 2: One Radian.

Because the circumference of a circle of radius R is $2\pi R$, there are $2\pi = 2 \times 3.14159$ radians in a full circle, that is, 2π radians is 360 degrees. If a roller is rotating through, say, $1/3$ radian every quarter inch (every observation), this would be called the frequency and could also be expressed as $4/3$ radian per inch. The roller would make one complete revolution every 6π (6×3.14159) observations, that is, every 18.8 quarter inches. Thus its circumference would be $6\pi / 4 = 4.7$ inches and its radius would be $3/4$ inch using our circumference = $2\pi R$ formula. The period would be the circumference, 1.5π inches or 6π observations. The horizontal axis of a periodogram plot represents the frequency of oscillation of the data in radians per observation (radians per quarter inch). The Y axis represents the importance of the frequency in determining the variation in the data. A large Y axis value (periodogram ordinate) at a given frequency would indicate an important cycle.

Look at a wheel with a mark on its edge. Glance back one second later and you see the mark is 45 degrees ($= \pi/4$ radians = $1/8$ revolution) counterclockwise from its starting position. You infer that the wheel is rotating at $1/8$ of a revolution per second counterclockwise. But is that true? Maybe it is rotating at $7/8$ revolution per second clockwise or $5/8$ revolutions counterclockwise. It follows that discrete observations like these wheel mark locations will indicate a unique angle only if frequencies are limited to an interval π radians (180 degrees). The standard choice is the interval from 0 to π (0 to 180 degrees). The theoretical spectrum is a symmetric function on the interval $[-\pi, \pi]$ so just showing half of it relays all the information. The area under the theoretical spectrum equals the variance of the time series, again showing how the variance has been broken down into frequency components of differing magnitudes.

With this introduction, we are ready to see if our data has cycles. None are apparent from the plot, but the periodogram appears in Figure 3 and tells quite a different story.

```
proc spectra data=fabric p s adjmean out=outspec; var Y;
  weights 1 2 3 4 5 4 3 2 1;
proc gplot data=outspec; plot P_01*freq S_01*freq;
  symbol1 v=diamond i=needle c=green;
```

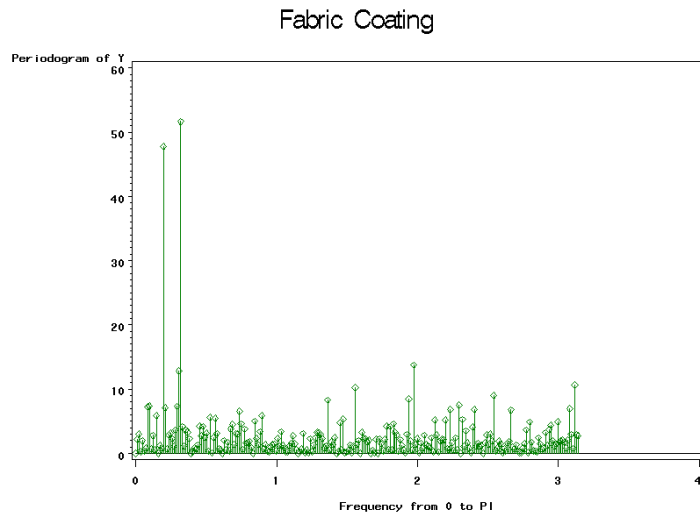


Figure 3: Fabric coating periodogram.

Note the use of the option “adjmean” which subtracts the series mean before analysis. This is typically done, for if not, there will appear a frequency 0 periodogram value equal to $n\bar{y}^2$. This would usually dominate any plot to the extent that the other values would appear to be right on the horizontal axis, giving no useful visual interpretation. The frequencies and periods (in terms of observations) are available in the output dataset OUTSPEC. Recall that frequency = 2π / period. The periodogram ordinates have variable name P_01 while S_01 is the name for the spectral density estimates, to be discussed later.

```
proc print data=outspec; where P_01 > 20;
Obs      FREQ      PERIOD      P_01      S_01
17      0.19635    32.0000    47.8072    0.92743
27      0.31907    19.6923    51.6533    1.17922
```

We have discovered that

- (1) Two of our rollers are likely out of line.
- (2) One roller that is out of line would rotate once in $2(3.14159)/.19635 = 32$ observations. With 4 observations per inch, that roller would have circumference 8 inches.
- (3) A roller of circumference 19.69 quarter inches (4.9 inches) is also indicated, but that is not one of our known circumferences. Perhaps it is really the 5 inch one.

BACKGROUND MATH

We see that this “periodogram” can discover hidden cycles in data. What exactly is a “periodogram”? As with many statistical concepts, the answer involves regression ideas. We need a small trigonometry refresher to start with. Observations over time following a sine wave are represented mathematically as

$$Y_t = A \sin(\omega t + C)$$

where A is the amplitude, ω the frequency in radians and C the phase shift. If C is 0, then the sine wave starts at $Y=0$, rises to $Y=A$, back to $Y=0$, down to $Y=-A$, and back to 0 as time t goes from 0 to $2\pi/\omega$. The wave repeats itself every $2\pi/\omega$ time periods (recall, frequency = 2π /period). A nonzero C means that at time 0 we have $\sin(C)$, not $\sin(0)=0$. Hence the name “phase shift.”

From trigonometry, we find for angles D and E the formula

$$\sin(D + E) = \sin(D)\cos(E) + \cos(D)\sin(E)$$

and applying this to our $A \sin(\omega t + C)$ we have $[A \cos(C)] \sin(\omega t) + [A \sin(C)] \cos(\omega t)$. Thus if we know the frequency ω and create variables

$$st = \sin(\omega t); \quad ct = \cos(\omega t);$$

For the fabric data, this suggests regressing Y on st and ct to estimate $[A \cos(C)]$ and $[A \sin(C)]$ from which estimates of A and C can be derived if desired. The following implements this idea:

```
data waves; pi=4*atan(1);
set fabric;
s5 = sin(2*pi*location/5); c5 = cos(2*pi*location/5);
s6 = sin(2*pi*location/6); c6 = cos(2*pi*location/6);
s8 = sin(2*pi*location/8); c8 = cos(2*pi*location/8);
s11 = sin(2*pi*location/11); c11 = cos(2*pi*location/11);
s16 = sin(2*pi*location/16); c16 = cos(2*pi*location/16);
s20 = sin(2*pi*location/20); c20 = cos(2*pi*location/20);
proc reg; model Y = s5--c20/ss1 ss2;
T_5: test s5=0, c5=0; T_6: test s6=0, c6=0;
T_8: test s8=0, c8=0; T_11: test s11=0, c11=0;
T_16: test s16=0, c16=0; T_20: test s20=0, c20=0;
```

The TEST statement results are

	5 inch	6 inch	8 inch	11 inch	16 inch	20 inch
$F_{499}^2 =$	37.0	0.22	20.08	3.27	2.72	1.36
Pr > F =	<.0001	.8018	<.0001	.0387	.0670	.2586

Knowing the potential contributing frequencies makes this problem fairly straightforward. We see that the 5 and 8 inch rollers need attention with some suggestion that the 11 inch one may also be slightly misaligned. Clearly the 5 and 8 inch adjustments have the most potential for making the coating uniform. PROC REG delivers an estimate of the standard deviation that would remain if all the rollers were perfectly aligned (Root MSE = 1.01354). In the PROC REG output, the Type I and Type II sums of squares are not the same, so our sine and cosine columns are not orthogonal to each other.

PERIDOGRAM CALCULATION

The situation is more challenging when we do not have a predetermined idea as to the important frequencies. As argued above, we can restrict ourselves to the frequencies in $[0, \pi]$, i.e. 0 to 180 degrees so with nothing more to go on, and realizing that each frequency requires 2 degrees of freedom, we might as well divide the interval into $(n-1)/2$ equally spaced frequencies if n is odd and $n/2$ (but with the last having only the cosine term) if n is even. These frequencies $\omega_j = 2\pi j/n$ are called "Fourier frequencies" and will deliver a set of $(n-1)/2$ sine and $(n-1)/2$ cosine columns if n is odd, $(n-2)/2$ sine and $n/2$ cosine terms if n is even. Adding the usual column of 1s, we have n observations on Y and n columns in our X matrix so the regression will fit perfectly. The sum of squares (Type I = Type II: orthogonal!) for each frequency is the periodogram ordinate for that frequency. It is for this reason that the plot of ordinates versus frequency shows which frequencies are important in "explaining" the variation in Y .

When n is odd, the Fourier frequencies give $n-1$ sine and cosine columns that are orthogonal and each has sum of squares $n/2$ except at frequency 0 (the mean) and, if n is even, at π . Apart from these frequencies, the regression $X'X$ matrix is $n/2$ times an identity. The regression coefficients of the sine and cosine columns, say A_j and B_j , at the j^{th} Fourier frequency can be inserted into the usual formula for regression sum of squares, $\hat{\beta}' X' X \hat{\beta}$ which becomes $(n/2) \sum_1^{(n-1)/2} (A_j^2 + B_j^2) + n\bar{Y}^2$ for n odd so that there is a sum of squares (periodogram value, periodogram ordinate) $I_n(\omega_j) = (n/2)(A_j^2 + B_j^2)$ for each Fourier frequency ω_j .

To illustrate, Figure 4 displays some columns of the X matrix graphically for $n=36$. The column of 1s is represented by a vertical line of 36 dots. The next two "columns" have dots at the sine and cosine of $\omega_1 = 2\pi/36$ so they go through 1 cycle in 36 observations, a period of 36 observations. The next pair are sine and cosine columns at $2\pi(2/36)$ and the last pair at $2\pi(3/36)$. The last two pairs of columns run through 2 and 3 cycles as the first pair runs through one. Thus any linear combination of these 6 columns will have period 36 but will not have the shape of a single sine wave. Such a wave is said to have fundamental frequency $2\pi/36$ with the other two frequencies being called "harmonics." For example, the note middle C on a piano has the same fundamental frequency as middle C on a trombone or other instrument. It is the wave shape produced by the harmonic frequencies that distinguishes the sounds on different instruments. Loudness is determined by the amplitude of the sound waves.

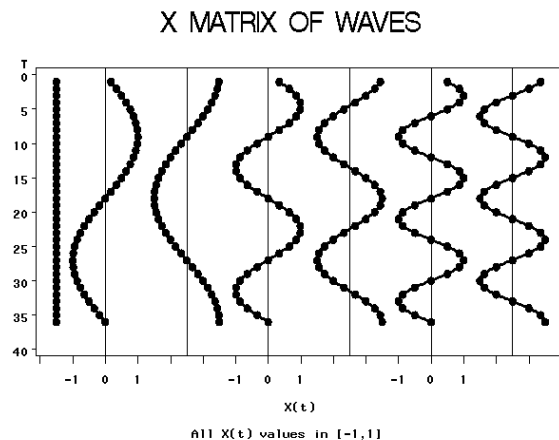


Figure 4: The X matrix.

SOME INTERPRETATION DETAILS

Our fabric coating dataset has 512 observations (128 inches of fabric) and $n/2 = 256$ is an exact multiple of 32. That means that increments of $1/256$ of a full revolution will exactly hit $1/32$ of a revolution, namely $8(1/256) = 1/32$ so that the frequency associated with period 32 observations (8 inches) will be exactly a Fourier frequency. Not so with period 20 observations (5 inches). There is no integer j such that $j/256 = 1/20$. Only $j = 12.8$ would work. In this case the variation associated with period 20 will "leak out" into neighboring periodogram ordinates, that is, we will get a spike or two at the very closest frequencies to

the one actually causing the fluctuations. This explains the finding of an effect at period 19.6923 in the periodogram of the fabric coating data.

A cycle of period 2 observations is the fastest oscillation we can detect with this technology. That frequency (π radians per observation) is called the “Nyquist frequency.” For example if we had a misaligned roller of circumference $\frac{1}{4}$ inch, that would be at the same place in its rotation on every observation and we could not detect its effect. A roller of circumference $\frac{5}{12}$ of an inch would make $\frac{3}{5}$ rotation or $\frac{6\pi}{5}$ radians per observation. Thus it would appear to be rotating $\frac{2}{5}$ rotation, that is $\frac{4\pi}{5}$ radians, backward per observation if we observe every quarter inch. We would say that the true frequency $\frac{6\pi}{5}$ is aliased with frequency $\frac{4\pi}{5}$. By measuring every quarter inch, we have more than two measurements around the circumference of any of our rollers and hence we have no aliasing problems from the rollers.

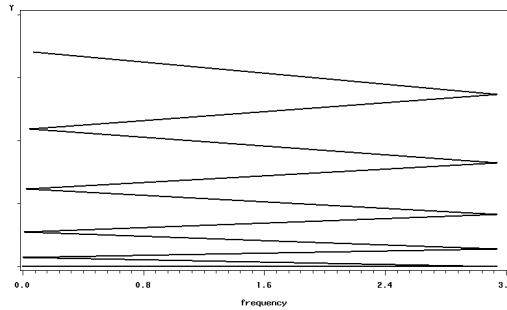


Figure 5: The Folding (Nyquist) Frequency and Aliases.

Imagine an axis of frequencies from $-\infty$ to ∞ . Mark the axis at 0 and every multiple of π . Leaving the interval $[0, \pi]$ where it is, fold the right side of the axis over at π so that 2π now rests on top of 0. Fold again (back to the right) at 2π , left at 3π etc. in accordion fashion. Do the same for the negative frequencies. For each frequency in $[0, \pi]$, for example $\frac{4\pi}{5}$, the frequencies that now lie above it, like $\frac{6\pi}{5}$ are its aliases. For this reason, the Nyquist frequency π is also called the “folding frequency.”

THE SPECTRUM

Is the periodogram an estimate of something? If so, what and how is it helpful to us? The answer lies in the so-called “spectrum.” You may recall that a prism will divide white light into bands of color – the color spectrum. The different frequencies of light waves contribute equally to white light whereas colored light would have some frequencies contributing more than others. In time series, the spectrum or spectral density is related to the covariance function and is defined in terms of angular frequency ω (in radians) as

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \cos(\omega h)$$

where $\gamma(h) = \text{Cov}(Y_t, Y_{t+h})$. For an uncorrelated, constant variance sequence we would have $\gamma(h) = \sigma^2$ when $h=0$ and $\gamma(h) = 0$ otherwise so the sum just boils down to $f(\omega) = \gamma(0) = \sigma^2/(2\pi)$, a constant function of ω . This constant $\sigma^2/(2\pi)$ plotted against ω on the interval $[-\pi, \pi]$ would be a horizontal line segment indicating equal contributions for all frequencies. The area under this function on that interval is the series variance σ^2 . In analogy to white light, such a series (uncorrelated, mean 0, constant variance) is called “white noise.” In fact any spectrum is viewed as breaking the variance of the time series into frequency components. For example the autoregressive order 1 process with autoregressive parameter ρ , in which $\gamma(h) = \rho^{|h|} [\sigma^2/(1-\rho^2)]$, has spectral density $\sigma^2/(2\pi(1+\rho^2-2\rho\cos(\omega)))$ and we see that when $\rho>0$, this function is highest at frequency 0 (more contributions from long waves) while $\rho<0$ gives smallest value at frequency 0 and hence more contributions from fast oscillating waves.

The periodogram divided by 4π estimates the spectral density. In PROC SPECTRA, P_0j denotes the periodogram and S_0j the spectral density estimate of the jth variable. Like the white noise and autoregressive spectra above, the spectra of all ARMA models are smooth functions of frequency and it is thus customary to average neighboring frequencies when computing S_0j. The WEIGHT statement gives the weighting scheme, for example, when estimating S_0j at the ith Fourier frequency, WEIGHT 1 2 3 2 1 gives triangular weights 1/9 to the i-2 and i+2 frequencies, 2/9 to i-1 and i+1 and 3/9 to the target ith frequency periodogram ordinate P_0j. The weighted average is divided by 4π to get S_0j.

To illustrate the idea of the autoregressive model, we try two plots using 13 triangular weights:

```

Data AR_1;
do t = 1 to 512; e= normal(123);
  Y = .8*Y + 2*e ; if Y=. then Y=0;
  X = -.8*X + 2*e; if X=. then X=0; drop e;
output; end;
proc spectra P S adjmean out=arspec; Var Y X;
weight 1 2 3 4 5 6 7 6 5 4 3 2 1;
data arspec; set arspec;
Ytheory = 4/(2*3.14159*(1.64 - 1.6*cos(freq)));
Xtheory = 4/(2*3.14159*(1.64 + 1.6*cos(freq)));
label S_01 = "Y (.8)";
label S_02 = "X (-.8)";
proc gplot; plot (S_01 Ytheory)*freq/overlay;
symbol1 v=dot i=none c=red h=.6;
symbol2 v=none i=join c=black w=3;
plot (S_02 Xtheory)*freq/overlay; run;

```

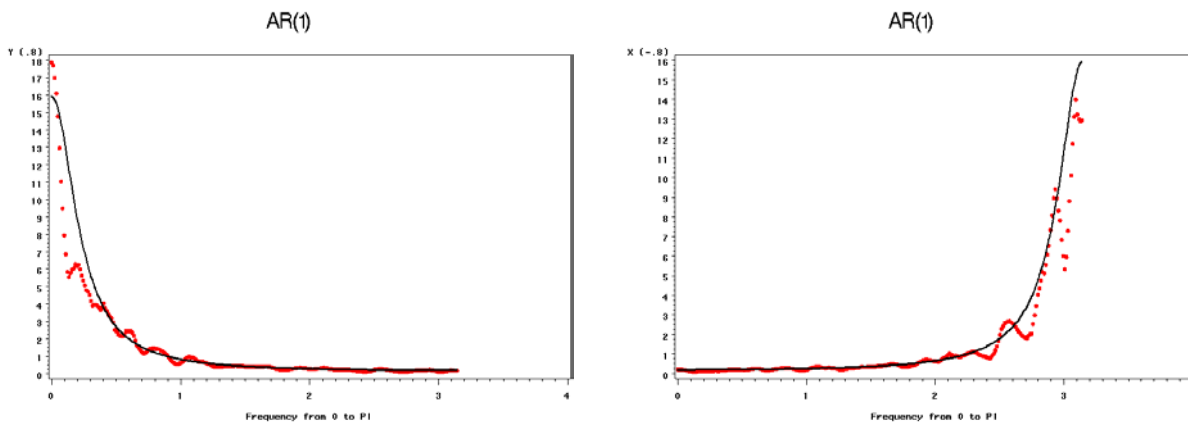


Figure 6: AR(1) spectra, .8 autocorrelation (left panel) versus -.8 (right panel)

The theoretical spectrum in black is well estimated by the smoothed periodogram S (the dots). By taking a moving average, autocorrelation in the estimates has been induced and overestimates tend to be surrounded by overestimates and likewise for underestimates. The variability is a function of the level.

For our roller data, here is the spectral density plot using 9 triangular weights:

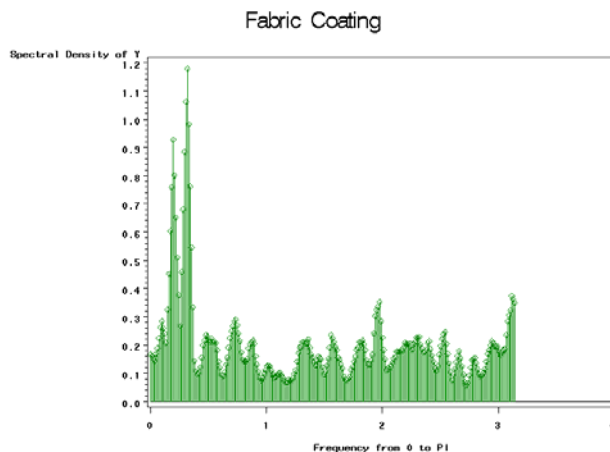


Figure 7: Smoothed Spectrum for Fabric Coating.

Other than the two spikes, the relatively flat spectrum is consistent with white noise. We appear to have two sinusoids buried in white noise. The down side to smoothing is that the effect of a pure sinusoid is spread across several frequencies so the spikes are not as sharp as before. The upside is that the smooth part is smoother. It is often the case that several sets of weights are tried and the graphs studied.

MORE EXAMPLES.

EXAMPLE 1: PLANT ENZYME ACTIVITY

Data are detrended measures of plant enzyme activity provided by Chiu-Yueh Hung, Dept. of Genetics, NC State U. recorded every 4 hours over 5 days. The researcher suspected a 12 hour cycle so frequency $2\pi/3$ per observation (2 cycles every 6 observations = 2 cycles per 24 hours) is of interest. Creating a sine and cosine at this frequency and regressing in PROC REG gives $F_{27}^2 = 7.09$ and $\text{Pr} > F$ is 0.0034 so there is evidence of a cycle.

Had we not anticipated this cycle, a periodogram might have been used. For such a case, Fisher proposed a test statistic, kappa for the null hypothesis that the series is just white noise. For $m-1$ calculated 2 degree of freedom ordinates, kappa is $(m-1)$ times the largest periodogram value divided by the sum of all $m-1$ values. Thus Fisher's test is constructed to detect a sine wave buried in white noise. Fuller (1996) discusses this test along with another test for white noise, Bartlett's Kolmogorov Smirnov test which uses a normalized cumulative sum of periodogram ordinates. This test is aimed at detecting a more general alternative and would hence be expected to have less power than kappa when the truth is a sine wave buried in white noise. Fuller's tables indicate a kappa value 4.877 is required for significance at the 5% level and 4.385 at the 10% level. Using

```
proc spectra whitetest;
```

We get both tests. The output appears as follows:

```
                ENZYME ACTIVITY
                The SPECTRA Procedure
                Test for White Noise for Variable Y

                M-1              14
                Max(P(*))       11932.89
                Sum(P(*))       34590.36

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

                Kappa          4.829682

                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          0.255984
                Approximate P-Value     0.3180
```

There are some good lessons in statistics here. Kappa is significant at the 10% level and almost, but not quite, at the 5% level. The more general Bartlett test would only be significant at the 32% level or more. We see the advantage of knowing something about the nature of the alternative. Also, if we are willing to declare knowledge of the appropriate frequency, we expect to gain even more power and indeed (recall) our regression F test had p-value 0.0034, quite a difference in conclusions!

A graph of the data with the regression estimated wave, and the periodogram are shown in the left and right panels of Figure 8.

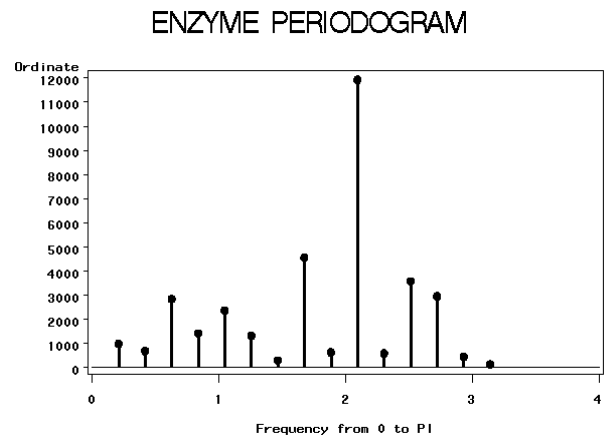
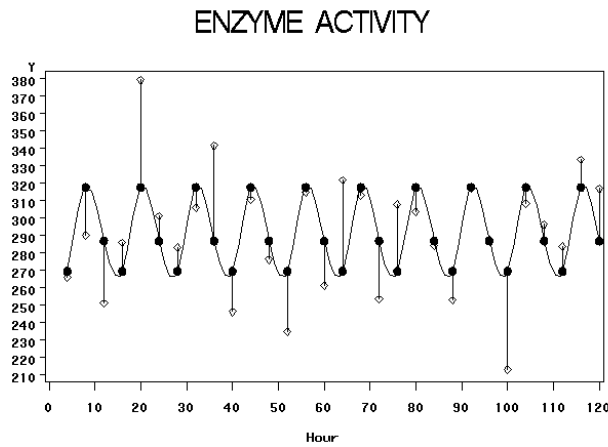


Figure 8: Plant Enzyme Activity.

EXAMPLE 2: NORTH CAROLINA RETAIL SALES.

The left panel of Figure 9 is a graph of monthly retail sales in North Carolina (log transformed). Overlaid on the data are predictions from two least squares regression models. On the right are periodograms

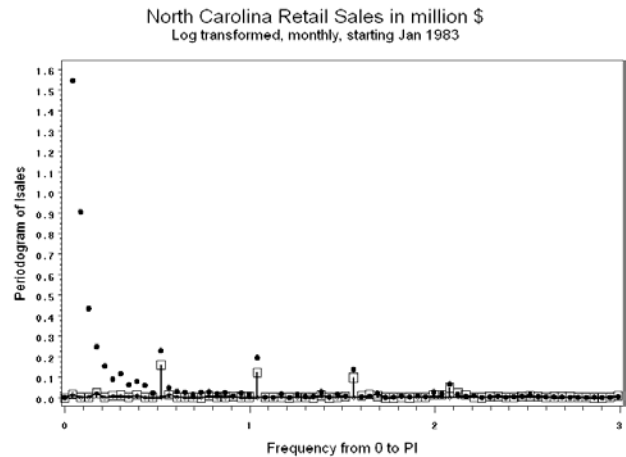
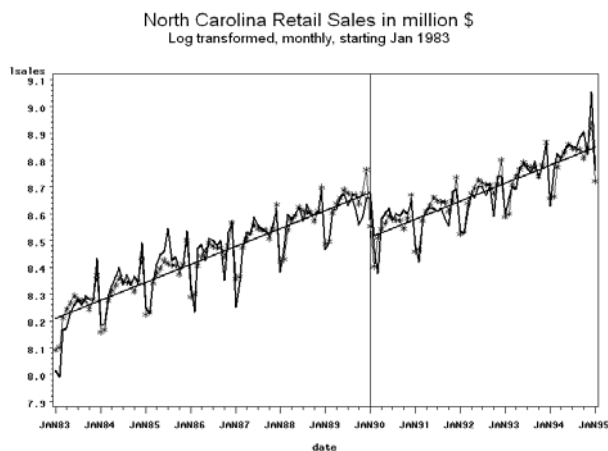


Figure 9: North Carolina Retail Sales.

The data are displayed as a thick black line in the top left plot. The periodogram of the data is shown as dots in the top right panel. Note the exceptionally high periodogram values at low frequencies. This comes from the trend in the data. Because periodogram analysis explains everything in terms of waves, an upward trend shows up as a very long (low frequency) wave. Also note the dots at multiples of $2\pi/12$ suggesting a strong seasonal pattern in the data. There appears to be a shift around January of 1990. A variable that is 0 up through that date and 1 thereafter, and a simple linear trend variable are used to predict the sales variable. The resulting predicting function is shown as two solid lines in the top left panel with a shift downward from one line to the next after January of 1990. The boxes atop needles in the right panel are the periodogram values for the residuals from this regression. We see how nicely the regression has removed those low frequency components. The seasonality remains as elevated boxes at multiples of $\omega=2\pi/12$. Three or four of these seem important, indicating that 6 or 8 sine and cosine terms would be needed to capture the seasonal effect. We could also capture the seasonal effect with 11 seasonal dummy variables (for example, a variable that is 1 only for January data and 0 otherwise) so we conclude the seasonal pattern is not very smooth. Of course that makes sense since retail sales tend to spike during the Christmas shopping season rather than slowly ascending and descending through the year like a sinusoid.

The top left panel also shows a second regression predicting function consisting of the same broken trend just described, plus 11 seasonal dummy variables. The predictions are asterisks on the graph connected by lines. The periodogram of the resulting residuals is shown as a thick line near the horizontal axis in the top right panel. Notice that the seasonality is completely gone now and little else has changed. That lower thick line runs through all the boxes except the seasonal ones. The periodogram looks quite flat, suggesting white noise.

Sometimes using the periodogram, as we have done here on sequences of model residuals, can be thought of as fine tuning the model. We have now removed what appears to be the majority of the variation in the data, resulting in a spectrum that looks almost flat, but let's take a closer look. First, we invoke the whitetest option on the original data, the trend adjusted data, and the trend plus seasonal adjusted data that has just been discussed. Here are the three Bartlett tests (original data, trend adjusted, trend and seasonal adjusted) and their p-values.

Test Statistic	0.639329
Approximate P-Value	<.0001
Test Statistic	0.19604
Approximate P-Value	0.0079
Test Statistic	0.153536
Approximate P-Value	0.0671

While the final model gives an insignificant test at the 0.05 level, it is close enough to significance to suggest some further checks. It would not be surprising to find autocorrelation in such data and we recall that positive autocorrelation would show up as low frequency components in our periodogram. The periodogram overlaid with a spline fit predictor from PROC GAM as shown in the left panel of Figure 10 indicates such a possibility.

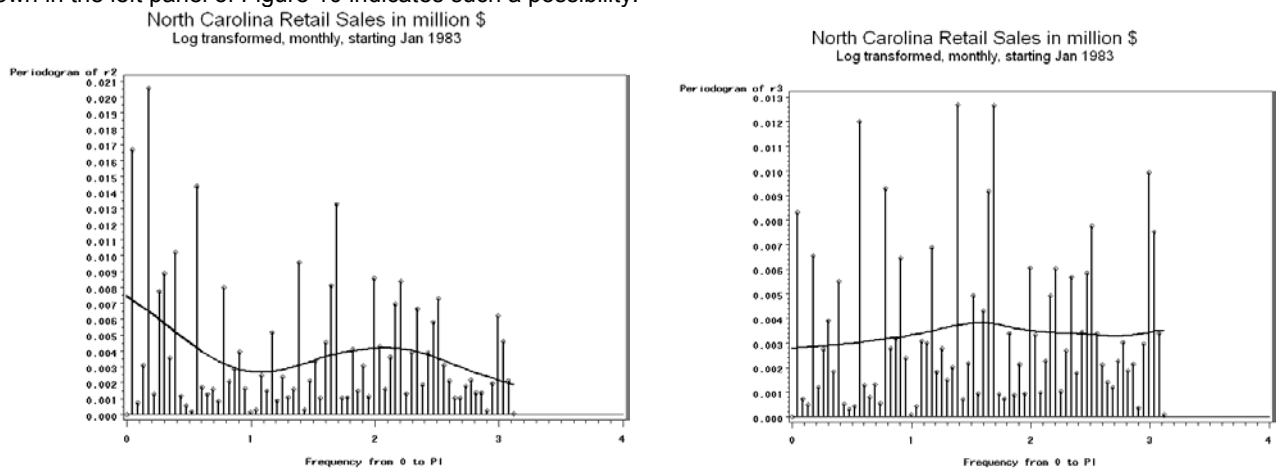


Figure 10 Residual Periodograms, NC Retail Sales.

Refitting the trend plus seasonal model in PROC AUTOREG with 3 lags for the error term results in residuals whose periodogram is quite flat as is shown in the right panel of Figure 10. Furthermore, the Bartlett test results on this periodogram are as follows, showing no evidence whatsoever against the white noise null hypothesis.

Test Statistic	0.051116
Approximate P-Value	0.9918

EXAMPLE 3: AIRLINE TICKET SALES

These are the famous international airline ticket sales used to illustrate seasonal time series used to illustrate seasonality in Box, Jenkins, and Reinsel (1994). Figure 11 is a SAS template with the top graph being the data (log transformed international airline ticket sales). We see a trend and strong seasonal pattern. In the second row of the display we see the periodogram of the original data with large values near frequency 0 as would be expected in a trending series like this. The graph to its right is the periodogram of residuals from linear trend. The code to generate the periodogram ordinates P_01 and frequencies FREQ in a dataset called detrended is listed below.

```
proc reg noprint data=air; model lair = date;
  output out=out1 residual = detrend;
proc spectra p data=out1 out=detrended;
var detrend;
```

On the left in the last row of graphs is the periodogram of the residuals from a model that includes a linear trend and seasonal dummy variables. The periodogram values at the seasonal frequencies are exactly 0 for such residuals. The seasonal dummy variables give the same predictions as a full complement of sinusoids and their harmonics at the seasonal (period 12)

frequency and because we have an exact number of years (144 observations) these are Fourier frequencies. There does still appear to be a fair amount of activity at the low end of the periodogram. We might think of this as arising from an autocorrelated string of error terms or it might indicate that the trend was not sufficiently flexible (for example we could try adding a quadratic term). Note the drastically different scales on the vertical axes of these plots.

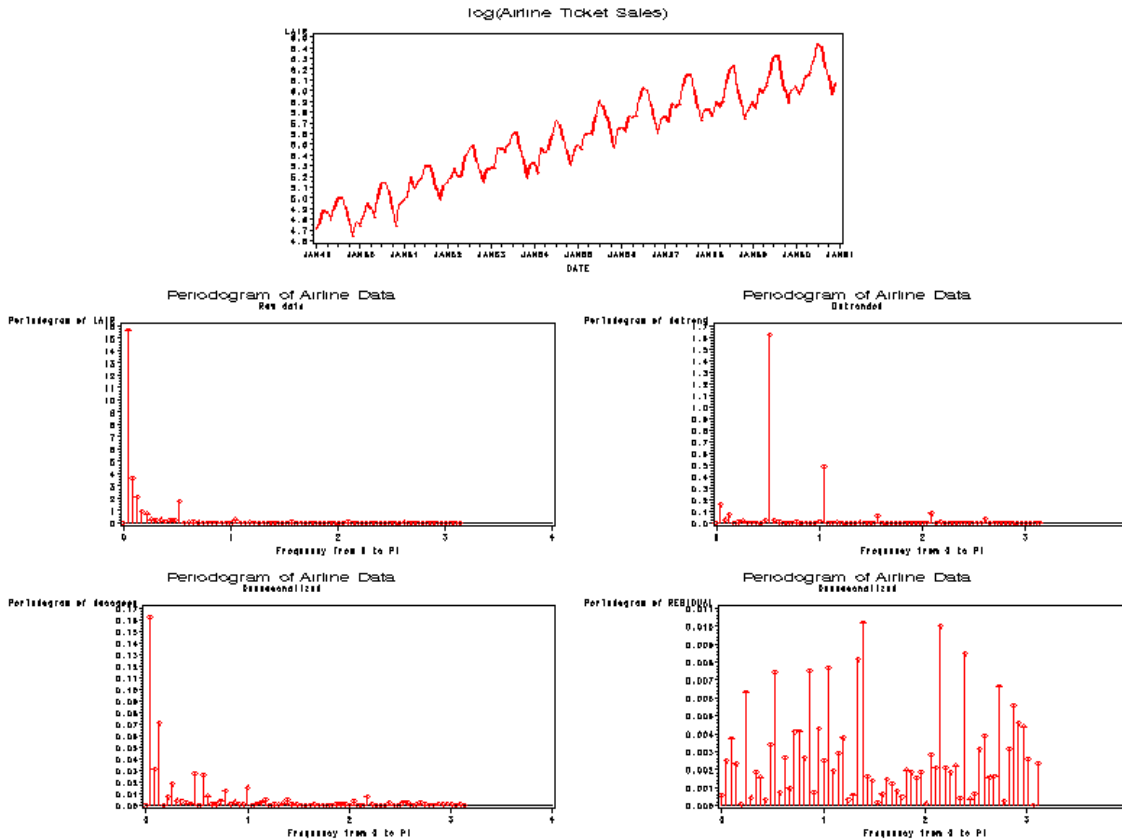


Figure 11: Airline Data.

Thinking about flexible models might lead one to consider a model with a more dynamic trend and seasonality. The model that has come to be known as the “airline model” because of this example is a time series model that allows a slowly changing seasonal and trend pattern. The model involves the “backshift” operator B . This operator is defined by $B(Y_t) = Y_{t-1}$. The model is

$$(1-B)(1-B^{12})Y_t = (1-\theta_1 B)(1-\theta_2 B^{12})e_t$$

To explain this model, suppose we look at the closely related model

$$(1-B)(1-B^{12})Y_t = e_t$$

The meaning of a product of backshift operators like this is just what you get by multiplication, that is

$$Y_t = Y_{t-1} + (Y_{t-12} - Y_{t-13}) + e_t$$

Suppose we are forecasting, say, February of 2003 (time t). The forecast would be the predicted value, $Y_{t-1} + (Y_{t-12} - Y_{t-13})$ which says to take the January 2003 ($t-1$) value and add to it the January to February change from 2002. Likewise a two step ahead forecast would add to this forecast the February to March change from last year and we observe that this process essentially takes the pattern for the last 12 months and tacks it onto the last observation. This model ignores everything except what happened last year. Adding the “moving average” terms $(1-\theta_1 B)(1-\theta_2 B^{12})e_t$, where the θ coefficients are between 0 and 1, causes the forecast to be a weighted average of all past years with weights declining as powers of those coefficients as we move back in time.

Recall that the spectrum is defined as the Fourier cosine transform of the covariance function. When a series requires differencing like this, it is referred to as nonstationary and there is not theoretical autocovariance function and hence no theoretical spectrum. However a stationary autoregressive seasonal model, for example the simple model

$$Y_t - \mu = \rho(Y_{t-12} - \mu) + e_t$$

has theoretical spectral density $\sigma^2/(2\pi(1 + \rho^2 - 2\rho \cos(12\omega)))$ that has high values at the fundamental seasonal frequency $2\pi/12$ and its harmonics. There would, theoretically, be high values near those frequencies as well, that is, we would expect more of a hill than a single sharp spike at these frequencies. If the autoregressive coefficient becomes close to 1, the large values at the seasonal frequencies would be expected to be even more pronounced. The airline model is an expansion of this seasonal autoregressive model, but would be expected to have similar behavior in terms of large values associated with seasonal frequencies.

The airline model can be fit in PROC ARIMA using this code, which also outputs the residuals and plots the periodogram and the smoothed spectrum estimate.

```
proc arima data=air;
  i var=lair(1,12) nlag=36 outcov=outcov3;
  e q=(1)(12);
  f id=date interval=month out=outfore;
proc spectra p s whitetest data=outfore out=out3;
  where residual ne .;
  var residual; weights 1 2 3 4 5 4 3 2 1;
proc gplot; plot P_01*freq; run;
```

The periodogram plot is at the bottom right of Figure 11 and looks flat, indicating white noise. The results of the white noise test also show a good fit.

Test for White Noise for Variable RESIDUAL

M	=	65
Max(P(*))		0.010204
Sum(P(*))		0.181609

Fisher's Kappa: M*MAX(P(*)/SUM(P(*)))

Kappa 3.651964

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	0.089766
Approximate P-Value	0.6715

MULTIVARIATE SPECTRA

TRANSFER FUNCITONS

An interesting multivariate model in time series is the so-called "transfer function" in which an output time series Y_t is related to current and past values of some input variable $X_t, X_{t-1}, X_{t-2}, \dots$ through a linear model that can be represented in general as

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \dots + Z_t$$

Where the noise process Z_t satisfies some ARMA time series model. Some special cases are of interest. If all but a finite number of the β s are zero, then the model can be estimated as a regression with time series errors in PROC AUTOREG (if the error series is autoregressive) or PROC ARIMA (using the CROSSCOR and INPUT options). With an infinite set of β s the estimation clearly cannot be done that way. In that case it is typical to formulate a model for these β s such that all of the β s depend on just a few parameters. An example of this would be $\beta_j = Cp^j$ where C and ρ are to be estimated.

The cross-covariance between Y and X is defined as

$$\gamma_{XY}(h) = \text{Cov}(X_t, Y_{t+h})$$

and just as we defined the spectrum theoretically as the Fourier transform of the autocovariances, there is also a theoretical cross-spectrum given by the Fourier transform of the cross-covariances. The Fourier transform is actually a complex valued function given by

$$f_{XY}(\omega) = \sum_{h=-\infty}^{\infty} \gamma_{XY}(h) \exp(i\omega h)$$

where i has the unfortunate name “imaginary number” and represents a number which when squared is -1 . An alternative expression for $\exp(i\omega h)$ is $\cos(\omega h) + i \sin(\omega h)$. Because the autocovariance function is symmetric in h , the sine terms drop out in the sum so it was easier to describe the formula for the spectrum. The cross spectrum comes out in a “real” plus an “imaginary” part, namely

$$\sum_{h=-\infty}^{\infty} \gamma_{XY}(h) [\cos(\omega h) + i \sin(\omega h)] = \sum_{h=-\infty}^{\infty} \gamma_{XY}(h) \cos(\omega h) + i \sum_{h=-\infty}^{\infty} \gamma_{XY}(h) \sin(\omega h)$$

The first term in this sum is called the cospectrum, let’s call it $c(\omega)$ and the second is i times the quadrature spectrum $q(\omega)$.

PURE SINE WAVE EXAMPLE

To see how to interpret this, look at the output from this program:

```
Data a; pi = 4*atan(1);
do t = 1 to 107; Ct = cos(2*pi*t/15); St=sin(2*pi*t/15);
  X = 3*Ct + 4*St;
  Y = 5 + X + 2*X1 + 3*X2;
output; X2=X1; X1=X; end;
proc print data=a (obs=5); title "Bivariate Data";
proc gplot; plot (X Y)*t/overlay;
  symbol1 v=none i=spline c=black r=2;
  title "Wave run through a Transfer Function";
proc reg; model Y = St Ct;
proc spectra data=a adjmean P cross K out=out2; var X Y;
  data out2; set out2;
  if RP_01_02 >0.01 then do;
  angle = (180/3.14159)*ATAN(IP_01_02/RP_01_02);
  gain = sqrt(P_02/P_01);
  amp = 5*gain; end;
proc print data=out2(obs=12); run;
```

The regression output shows that Y , which depends on several lags of X , is a perfect linear combination of the sine and cosine variables st and ct that generated X .

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	42829	21414	Infity	<.0001
Error	102	0	0		
Corrected Total	104	42829			

Root MSE	0	R-Square	1.0000
Dependent Mean	5.00000	Adj R-Sq	1.0000
Coeff Var	0		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	5.00000	0	Infty	<.0001
St	1	28.46665	0	Infty	<.0001
Ct	1	2.33182	0	Infty	<.0001

The amplitude of the X wave is 5, the square root of the sum of squared coefficients 3 and 4 used in generating X from st and ct. The amplitude of a sine wave is half the distance from peak to trough and in Figure 12 below, the X wave is seen to fluctuate between -5 and 5. Likewise the amplitude of Y, also shown in Figure 12, is the square root of $28.4665^2 + 2.33182^2$ which is 28.5620. The transfer function has passed the period 15 wave from X into another period 15 wave in Y, but the amplitude has increased from 5 to 28.5620. That is an increase, or gain as it is called, of $28.5620/5 = 5.7124$. Recall that the idea of spectral analysis is to break series down into sinusoidal components. In this case there is only one, but if there were several, as in a real data example, each frequency would have its own amplitude increase or decrease.

Wave run through a Transfer Function

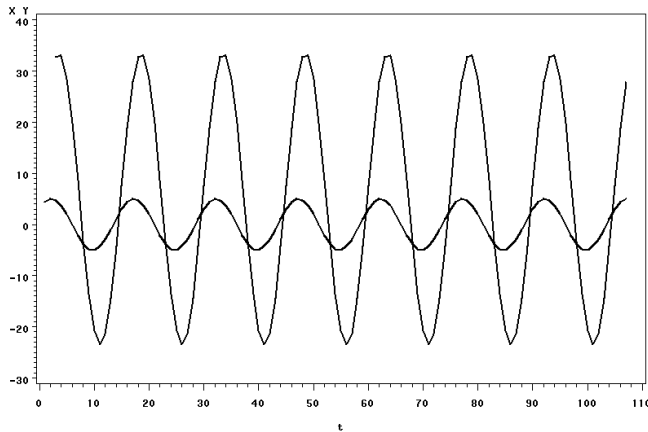


Figure 12: Sinusoidal Waves for Transfer Function

Looking at the graph carefully it is seen that, not only has the amplitude increased, but the peaks in Y occur just after those of X, that is, these two waves are not in phase. The difference in phase is referred to as the phase shift between X and Y at our frequency $2\pi/15$.

To compute this phase shift, recall that

$$A \sin(\omega t + s) = A \sin(\omega t) \cos(s) + A \cos(\omega t) \sin(s)$$

so the amplitude of the generic sinusoidal wave here is A and s represents the phase shift. If we regress this wave on $\sin(\omega t)$ and $\cos(\omega t)$ the coefficients are $A \cos(s)$ and $A \sin(s)$ and it is seen $A^2 \cos^2(s) + A^2 \sin^2(s) = A^2$, the amplitude squared. The ratio $[A \sin(s) / A \cos(s)]$ is $\tan(s)$ from which the arctangent function $\text{ATAN}(\)$ could then be used to recover s. This is the phase shift of a single wave relative to a pure sine wave. Computing s for both X and Y gives their phase shifts relative to the pure sine and the difference of these is then the phase shift of Y relative to X.

The following program asks for cross spectral quantities as these are called. The real and imaginary parts of the periodogram are computed and in a datastep, the arctangent function is used to compute the phase shift, which is converted to degrees.

FREQ	PERIOD	P_01	P_02	RP_01_02	IP_01_02	angle	gain	amp
0.00000	.	0.0	0.00	0.00	0.00	.	.	.
0.05984	105.000	0.0	0.00	0.00	0.00	.	.	.
0.11968	52.500	0.0	0.00	0.00	0.00	.	.	.
0.17952	35.000	0.0	0.00	0.00	0.00	.	.	.
0.23936	26.250	0.0	0.00	0.00	0.00	.	.	.
0.29920	21.000	0.0	0.00	0.00	0.00	.	.	.
0.35904	17.500	0.0	0.00	0.00	0.00	.	.	.
0.41888	15.000	1312.5	42828.86	6345.26	3993.82	32.1871	5.71240	28.5620
0.47872	13.125	0.0	0.00	0.00	0.00	.	.	.
0.53856	11.667	0.0	0.00	0.00	0.00	.	.	.
0.59840	10.500	0.0	0.00	0.00	0.00	.	.	.
0.65824	9.545	0.0	0.00	0.00	0.00	.	.	.

The periodograms show that both X (P_01) and Y (P_02) consist only of a period 15 wave. The phase angle is 32.19 degrees which seems reasonable from the graph. The gain 5.71240 and the resulting amplitude 28.562 of Y match our computations.

WHITE NOISE INPUT

Of course with real data, the input series would typically be a random process. In the program just above, if we change the definition of X to $X=5*\text{NORMAL}(123)$ then X still has mean 0, but is white noise now so its spectrum is flat (equal contributions from all frequencies) rather than having the fluctuations in X all at one frequency. Making that change we have a new X and Y series related by the same transfer function as before. It is typical and desirable to use local weighted averages in this more realistic case when computing the spectral quantities. The data can be analyzed using

```
proc spectra data=a adjmean P S cross Ph A K out=out2;
var X Y;
```

The new options here are Ph (phase spectrum estimates) A (cross amplitude estimates) and K (squared coherency estimates). The cross amplitude spectrum was the amplitude of the Y sine wave in our previous example. In general it is the amount of the Y amplitude that is attributable to X (at each frequency). Our previous transfer function examples so far have had no noise and so the total amplitude of Y is just the amount contributed by X. If we think of the relationship between Y and X as a regression in the frequency domain, the squared coherency is interpretable as the R-square at each frequency. Simply put, it is a number between 0 and 1 for each frequency, measuring the strength of the relationship between Y and X. Not surprisingly, the squared coherency is 1 at all frequencies in our noiseless transfer functions (sinusoidal or white noise inputs).

Plotting the phase spectrum (PH_01_02 in the output dataset out2) against frequency gives Figure 13.

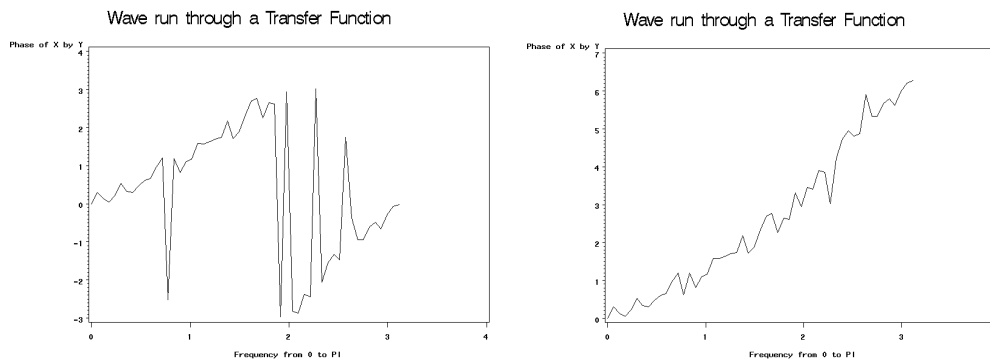


Figure 13: Phase Spectrum, Noiseless Transfer Function.

The seemingly radical jumps in the left panel are easily explained. Mathematically, since the phase spectrum is an arctangent function and the tangent of an angle s is indistinguishable from that of $s+j\pi$ for any integer j , one can add or subtract multiples of π to make the graph smoother. Making such shifts as in the right panel of Figure 13, or just ignoring the sudden shifts in the left panel graph, we see a roughly linear increase with slope about 1.5. The interpretation is that there is roughly a 1.5 time period delay between movements in X and the corresponding response in Y. This makes some sense since Y was generated from X by the line of code

$$Y = 5 + X + 2*X1 + 3*X2;$$

where the X1 and X2 were lags and the weights 1, 2, 3 would suggest an average “delay” somewhat over 1 day.

In the previous examples, Y_t is related to several past X values. An interesting thing happens when Y_t is a multiple of X_{t-d} plus noise. Here d need not be an integer. The slope of the phase spectrum in this case is an estimate of d . Thus in the case of a pure delay of time d , not necessarily an integer, the slope of the phase spectrum can be used to estimate the delay. Recall that with real data, there is also typically a noise component in the transfer function. Thus before interpreting the slope one might want to check the R-square measure. Recall that this is a measure interpretable as a frequency by frequency R-square and is referred to as Kappa-square.

Brocklebank and Dickey (2003) give an analysis of some streamflow data at two locations. The original flows are log transformed and differenced, resulting in Y (downstream) and X (upstream) series that are interpretable roughly as percentage

(proportion actually) changes in flow. They show how to estimate the transfer function in the time domain. The relationship between Y and X is

$$Y = \text{constant} + .4955 X_{t-1} + .2727 X_{t-2} + \text{Noise}$$

where the noise term has a time series (ARMA) structure. The X_t series also seemed to satisfy an ARMA time series model and the coefficients were estimated by the authors. If these estimated coefficients were the theoretical ones, then the covariance between Y_{t+h} and X_t could be computed using the formulas above as could the Fourier transform and hence the phase and amplitude spectra. One finds

$$\text{Cov}(Y_{t+h}, X_t) = \gamma_{XY}(h) = .4955 \gamma_{XX}(h-1) + .2727 \gamma_{XX}(h-2)$$

where $\gamma_{XX}(h)$ is the autocovariance of X at lag h which can be calculated as though the X model is the truth. Hence the Fourier transform of the cross covariance is related to the spectral density of X.

Brocklebank and Dickey (2003) discuss this computation and the resulting formulas in some detail. The formulas for estimating cross spectral quantities directly from the data, that is without fitting a time domain model, are given in Fuller (1996). With such real data, one would almost always want to get the smoothed estimates. In the code above the spectral density estimate S_01 is just the periodogram estimate P_01 divided by 2π whereas the inclusion of a weight statement would result in S_01 being a local weighted average of periodogram values. Figure 14 presents graphs of the phase (left panel) and squared coherency (right panel).

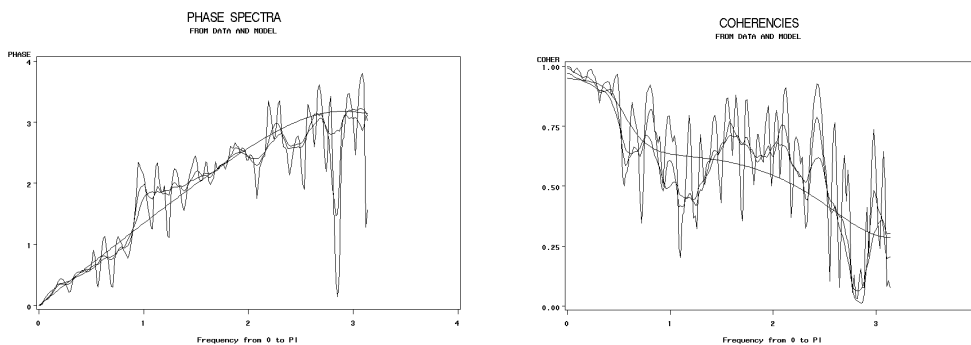


Figure 14 Phase and Coherency for River Data.

The very smooth graph in each panel is the theoretical plot for a series whose parameters equal the estimated time domain parameters of the fitted transfer function model. The three wiggly curves result from applying sets of 5, 13, and 21 triangular smoothing weights with wiggling becoming less as the number of weights increase. It is fairly common to try several sets of weights, and these weights are often called spectral “windows”. The coherency is very strong in the low frequencies. Long term fluctuations in upstream percentage changes in flow are very good predictors of the downstream changes. Fuller (1996) on page 393 gives a formula for a confidence interval for the phase spectrum at any given frequency when estimates use equal weights. Inspection of the formula shows that uncertainty in a phase estimate is directly related to the corresponding coherency in such a way that high coherencies give accurate phase estimates. In our example, the phase spectrum at low frequencies should be pretty accurate for this reason. The slope of the phase spectrum for frequency less than 0.4, a region where the coherency is exceptionally strong, could be estimated by regressing the phase spectrum values on frequency. This slope gives the estimated time delay in days between the two river gauging stations. For all three levels of smoothing, delays of around 1.3 days are found. Sine and cosine waves with frequencies less than 0.4 have periods greater than $2\pi/0.4$, about 16 days.

Finally, the gain of Kinston over Goldsboro represents an amplitude multiplier for each frequency. One would expect short term fluctuations in the day to day percentage change in flow at Goldsboro to smooth out by the time the river flows to Kinston. The gain should thus be quite a bit smaller than 1 at high frequencies. If these percentages rise and fall very slowly along a low frequency wave, one might expect most of that action to transfer downstream with a resulting gain near 1. There would be no reason here to expect a gain larger than 1 although in general that can happen. Obtaining large gains in sound waves, for example, is the purpose of the microphone and sound systems at concerts and sporting events. Figure 15 shows that these anticipated features are realized in both the model and data based gain calculations, however Figures 14 and 15 do seem to reveal some data features that are not completely reproduced by the model.

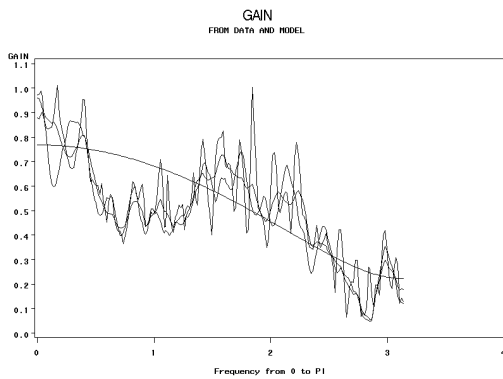


Figure 15 Gain for River Data

SUMMARY

This paper reviews the analysis of time series data using SAS PROC SPECTRA. This procedure allows one to decompose the variation in a time series into components corresponding to different periodicities. Results can be plotted to graphically reveal regular sinusoidal movements in a series which may or may not be anticipated from the nature of the observations. These visual impressions can be checked with statistical tests. Autocorrelated time series can be analyzed as well and it is seen that different autocorrelation structures induce a variety of long and short term movements in data. Smoothing of plots is often done to produce better estimates of the spectral density.

Multiple series can be studied in pairs. Available techniques allow estimation of the time delay between fluctuations in one series and fluctuations of the same frequency in another. A frequency based measure called coherency, or more carefully called squared coherency, corresponds to the R-square measure in regression, allowing the user to assess the strength of relationship at low (long run), medium, and high (short run) frequencies. A gain function is available to describe the amplification or attenuation of waves at varying frequencies.

REFERENCES

- Box, G.E.P., G. M. Jenkins and G. C. Reinsel (1994). Time Series Analysis: Forecasting and Control. Prentice Hall, Englewood, NJ.
- Brocklebank, J. C., and D. A. Dickey (2003) SAS for Forecasting Time Series 2ed. SAS Institute, Cary NC.
- Fuller, W. A. (1996). Introduction to Statistical Time Series. Wiley, NY.