Package name: Periodogram
Author: Nicolas Ronderos Pulido.
Date: November 25, 2013
Add-in type: Series object
Default proc name: periodogram
Default menu text: Periodogram

Spectral analysis

Spectral analysis aims to decompose a stationary (covariance stationary or weakly stationary) time series into an infinite sum of sine and cosine deterministic components, that helps to explain the variance of a series. The significant components explain economic cycles, seasonality and the unobserved components of a time series.

Unlike the time series analysis, the spectral analysis works in the frequency domain. The two analyzes are not mutually exclusive, so a signal component could be related to a lag in the time domain.

In one hand, the time domain Wold Theorem states that any stationary time series can be represented by an infinite sum of AR(p) and MA(q) terms. On the other hand, in the frequency domain, the Spectral Representation Theorem states that any stationary series can be represented by an infinite sum of sine and cosine signals —Hamilton (1994)—.

A stationary time series x_t could be expressed as a k term sum of cosine signals (1).

$$x_t = \sum_{i=1}^{k} [C_i Cos(\omega_i t + \phi_i)] + e_t \quad (1)$$

Where C_i is the amplitude, ω_i is the angular frequency given in radians, ϕ_i is the phase angle and e_t is the non-explained part of the series using the k terms. Following Montenegro (2009), the expression in (1) can be simplified using the trigonometric identity on (2).

$$Cos(\omega_i t + \phi_i) = Cos(\omega_i t) Cos\phi_i + Sin(\omega_i t) Sin\phi_i \quad (2)$$

Transforming the expression in (1) as:

$$x_t = \sum_{i=1}^{k} (A_i Cos\omega_i t + B_i\, Sin\omega_i t) + e_t \quad (3)$$

Where $A_i = C_i Cos\phi_i$ and $B_i = -C_i Sin\phi_i$. The equation in (3) could be seen as an OLS regression, where x_t is explained by the k signals on the right side of the equation. The matrix of explanatory variables is formed by sine and cosine variables with different frequencies, which have the useful property of orthogonality.

Assuming that we know ω_i, the coefficients by OLS would be:

$$\widehat{A_\iota} = \frac{2}{n}\sum_{t=1}^{n} x_t Cos(\omega_i t) \quad (4)$$

$$\widehat{B_\iota} = \frac{2}{n}\sum_{t=1}^{n} x_t Sin(\omega_i t) \quad (5)$$

In practice we do not know ω_i that could explain the variance of x_t. So we need to test a sequence of ω_i and register the values of $\widehat{A_\iota}$ and $\widehat{B_\iota}$ to see which frequencies are significant on x_t spectrum.

The Periodogram

The periodogram is the estimator of the area under the population spectrum of a series, which can be seen as a probability distribution function.

In general, a stochastic process would not have fixed frequencies that explain the variance of this process, then, the periodogram should not be interpreted as the portion of the variance that is explained with a frequency exactly equal to ω_i, it should rather be interpreted as frequencies near to ω_i that help to explain the variance.

The graph of $\sqrt{\widehat{A_\iota^2} + \widehat{B_\iota^2}}$ versus ω_i is known as the periodogram and it is denoted by $I_n(\omega)$. When a frequency is contained in x_t the value of the periodogram for that frequency will be different than zero. In the estimation of the periodogram the ω_i sequence is taken as follows for practical reasons.

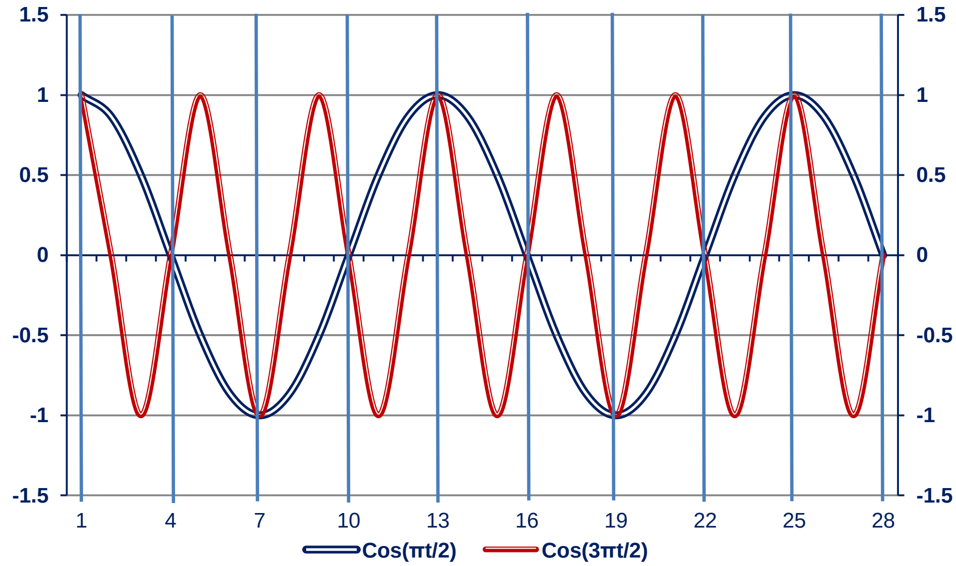
$$\omega_i = \frac{2\pi q_i}{n} \quad for \;\; q_i = 0,1,\dots,\frac{n}{2} \quad (6)$$

Therefore (6) equals to:

$$\omega_i = \left\{0, \frac{2\pi}{n}, \frac{4\pi}{n}, \dots, \pi\right\}$$

The frequency is in the range $0 < \omega_i < \pi$ because the other frequencies are subject to *aliasing* and also the periodogram is symmetric[1]. Aliasing is basically when a series generated with a sine or cosine function, with frequency bigger[2] than π and the data is observed only at integer values (for example *t*) the *observed data* will be exactly the same as a series generated with a frequency satisfying $0 < \omega_i < \pi$, so the sequence of frequencies that satisfy the last condition will be a generalization of all possible observed cycles. An example is showed in graph 1, where the blue lines highlight the observed matching points, note that the frequency of $Cos\left(\frac{3\pi t}{2}\right)$ is outside $0 < \omega_i < \pi$. Therefore, a series that is generated with a frequency outside the range will be observed as one inside the range.

Graph 1 —*Aliasing*—



As mentioned before, the spectral analysis is not independent from the time series analysis; this can be seen in (7).

$$I_n(\omega) = \psi \sum_{\tau=-(n-1)}^{n-1} \hat{R}(\tau) Cos\tau\omega \quad (7)$$

Where ψ denotes a constant of proportionality and $\hat{R}(\tau)$ is the autocovariance under de assumption of $Ex_t = 0$. For example, if the autocovariance shows a positive correlation, like in a AR(1) with positive coefficient, there will be a predominance of low frequencies in $I_n(\omega)$, and in AR(1) with negative coefficient,

[1] For more details see Hamilton (1994) chapter 6 pages 160-161.

[2] Or smaller tan $-\pi$ considering the symmetry of the periodogram

there will be a predominance of high frequencies in $I_n(\omega)$. From (7) it can be inferred that a white noise process has a flat spectrum, and the explained variance of the series by the frequencies will be distributed across the spectrum.

Sample Properties

As Hamilton (1994) mention, the periodogram is an unbiased estimator of the population spectrum, but it is inconsistent because we are estimating as many parameters as observations, therefore, the periodogram variance do not decreases as the sample size increases.

Spectral density

The tested frequencies that protrude in the periodogram are not exactly the frequencies that explain the variance of the series, rather are frequencies near to the significant ones that help to explain the variance. This assumption is the basis of the *nonparametric* or *kernel estimates*. It is also called *power spectral density* and it is denoted by $I_k(\omega_j)$, which follows the next expression.

$$I_k(\omega_j) = \sum_{m=-h}^{h} k(\omega_{j+m}) I_n(\omega) \quad (8)$$

Where h is the bandwidth, that indicates how many estimates from the periodogram will be averaged for calculate the spectral density. The k function is the Kernel and indicates the distribution of the weights. The Kernel function must satisfy the (9) condition.

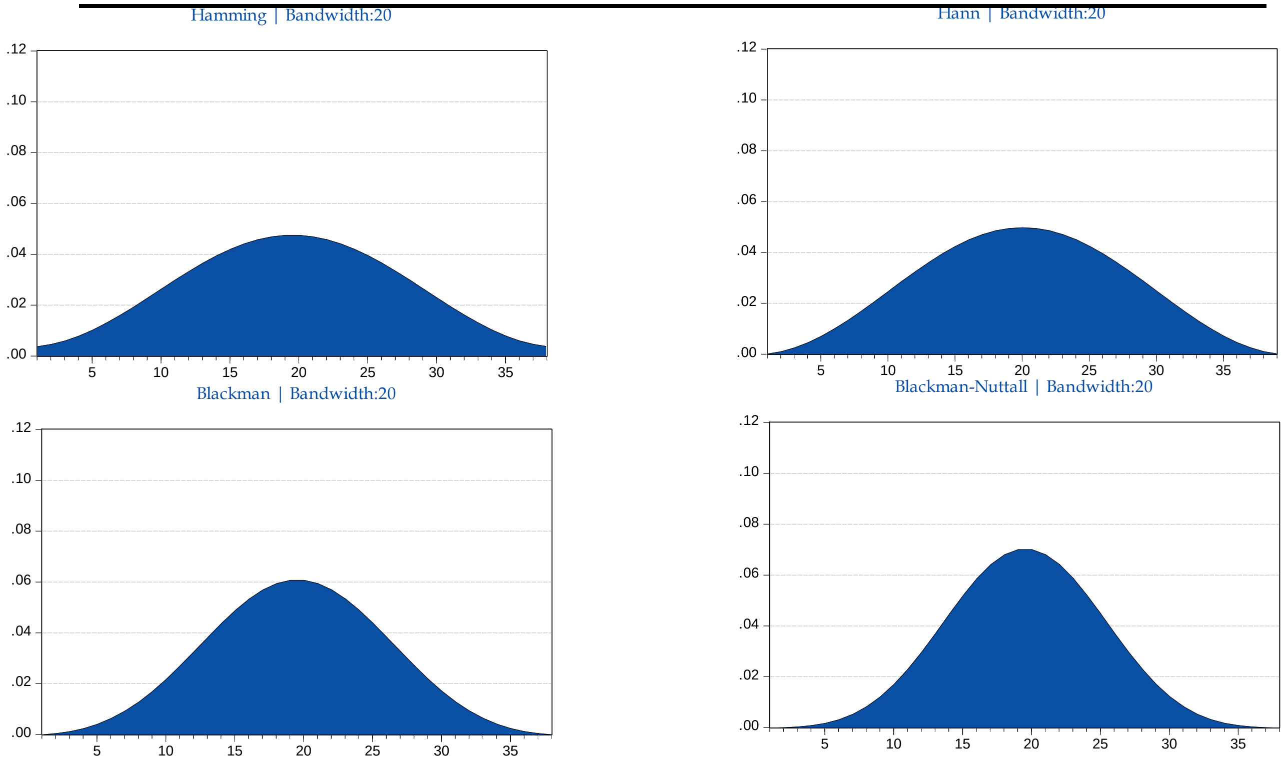
$$\sum_{m=-h}^{h} k(\omega_{j+m}) = 1 \qquad (9)$$

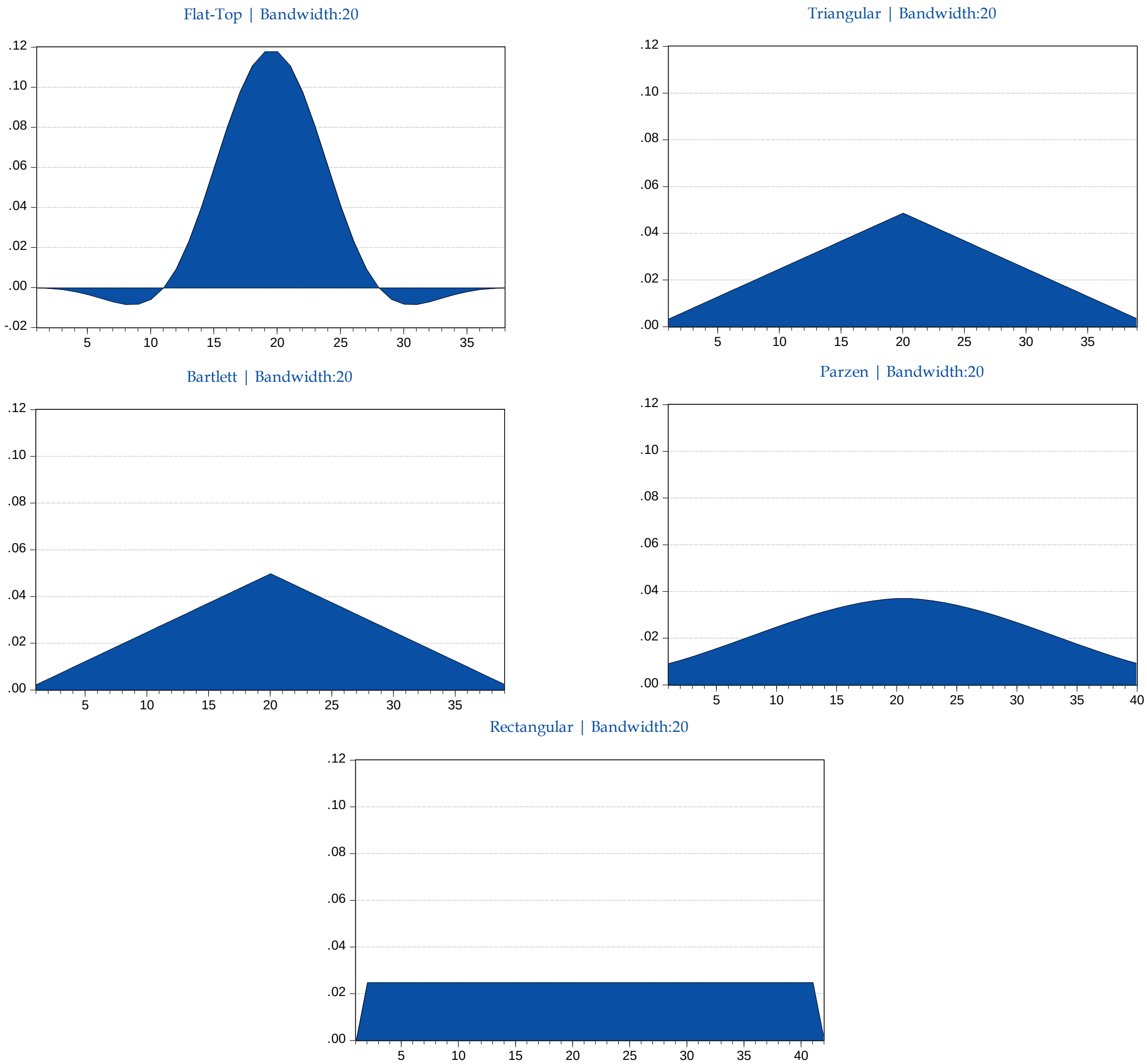
The Kernel windows used in the add-in will be described below with their functions (table), their distribution of weight (graph 2) and a short explanation for the bandwidth selection criteria. Most of the windows were taken from Stoica and Moses (2005).

The selection of the bandwidth in very important since a high value of it could introduce some bias; but a small value could not reduce enough the variance of the periodogram, therefore it is necessary and intermediate value. There is not a unique select criterion of the bandwidth. Hamilton (1994) suggests trying different values and on subjective judgment choosing the most plausible estimate. Another one is start with a low bandwidth and increase it until the periodogram is smoothed, because it is easier to identify visually smoothing than backwards. There are also optimal suggestions for the bandwidth for each type of Kernel, which in general vary with the sample size. For example, it is recommended use for the Bartlett $\sqrt[3]{t}$ and for the Parzen, Hamming and Hann$\sqrt[5]{t}$. Another general suggestions are $2\sqrt{t}$, $\sqrt{t}$ and $\frac{t}{40}$.

Name	**Function**
Hamming	$k(h) = 0.53836 - 0.46164\ cos\left(\frac{2\pi h}{M-1}\right)$
Hann	$k(h) = 0.5 - 0.5\ cos\left(\frac{2\pi h}{M}\right)$
Blackman	$k(h) = 0.42 - 0.5\ cos\left(\frac{2\pi h}{M-1}\right) + 0.08\ cos\left(\frac{4\pi h}{M-1}\right)$
Blackman-Nuttall	$k(h) = 0.36 - 0.49\ cos\left(\frac{2\pi h}{M-1}\right) + 0.14\ cos\left(\frac{4\pi h}{M-1}\right) - 0.01\ cos\left(\frac{6\pi h}{M-1}\right)$
Flat-Top	$k(h) = 1 - 1.9\ cos\left(\frac{2\pi h}{M-1}\right) + 1.29\ cos\left(\frac{4\pi h}{M-1}\right) - 0.39\ cos\left(\frac{6\pi h}{M-1}\right) + .03\ cos\left(\frac{6\pi h}{M-1}\right)$
Triangular	$k(h) = \frac{M+1}{2} - \left\lvert k - \frac{M}{2} \right\rvert$
Bartlett	$k(h) = \frac{M}{2} - \left\lvert k - \frac{M}{2} \right\rvert$
Parzen	$k(h) = \begin{cases} 1 - 6\left(\frac{h}{M}\right)^2 + 6\left(\frac{h}{M}\right)^3 & 0 \le h \le \frac{M}{2} \\ 2\left(1 - \frac{h}{M}\right)^3 & \frac{M}{2} \le h \le M \end{cases}$
Rectangular	Centered moving average

Graph 2 —*Kernel*—



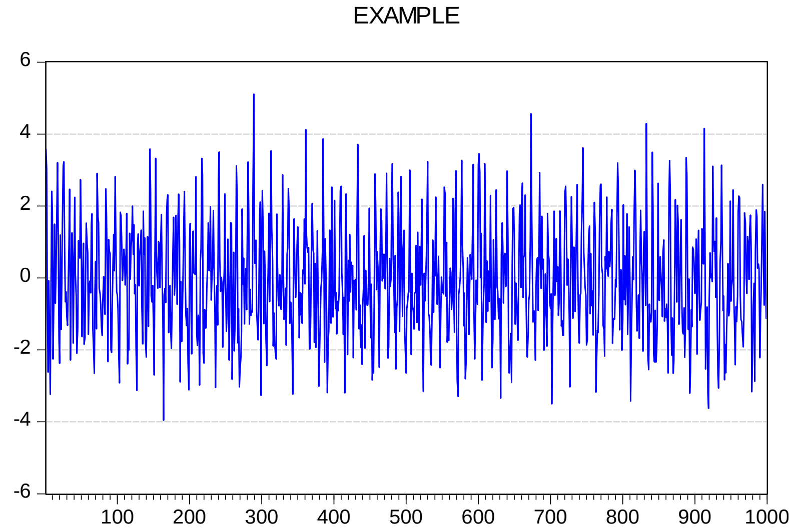


Example 1

The series in graph 3 was simulated with sine and cosine function with the command showed below for 1000 observations; in this case we know exactly how the data generator mechanism is.

*genr x=@cos(2*3.14159*@trend/8)+@cos(2*3.14159*@trend/12)+@sin(2*3.14159*@trend/4)+nrnd*

Graph 3



Given that a cosine or sine function has the structure in (10).

$$CSin(\omega t + \phi) \;\; or \;\; CCos(\omega t + \phi) \;\; (10)$$

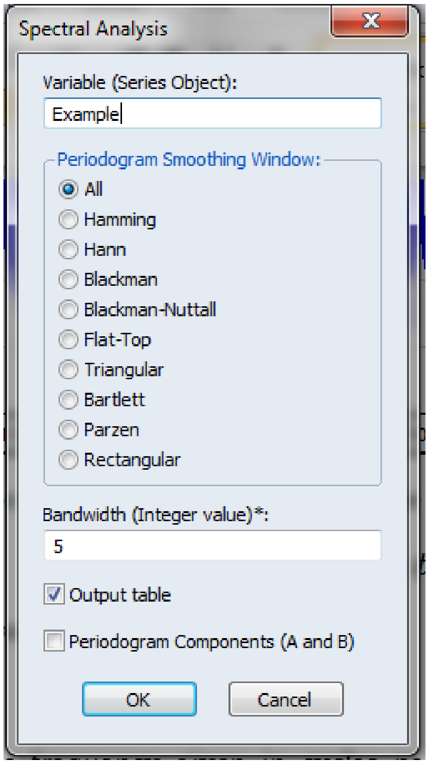
And ω could be expressed as:

$$\omega = 2\pi f$$

Where f express the frequency given in cycles per time unit. In the estimated periodogram of the series example it is expected that the frequencies taken from $2\pi f = \frac{2\pi}{8}$, $2\pi f = \frac{2\pi}{12}$ and $2\pi f = \frac{2\pi}{4}$ which are respectively $\frac{1}{8}$, $\frac{1}{12}$ and $\frac{1}{4}$ protrude in the spectrum of the series. Also, it is expected that across the periodogram there will be some frequencies that may appear to explain the variance of the series in small magnitude, which will be false, due to the white noise stochastic variable included in the simulation[3].

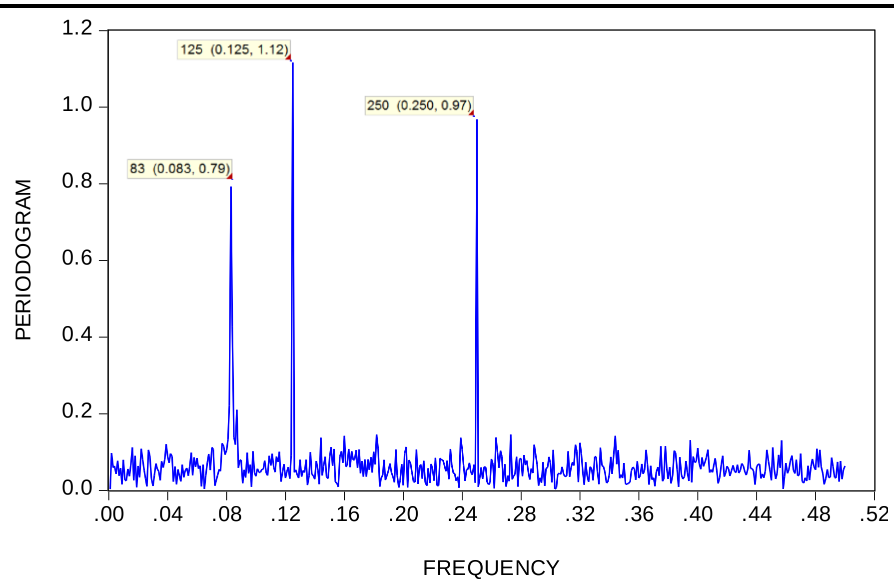
For estimating the periodogram you just have to click in the series object proc-add-inn-periodogram, and the following window will appear.

[3] $Where\ nrnd{\sim}N(0,1)$



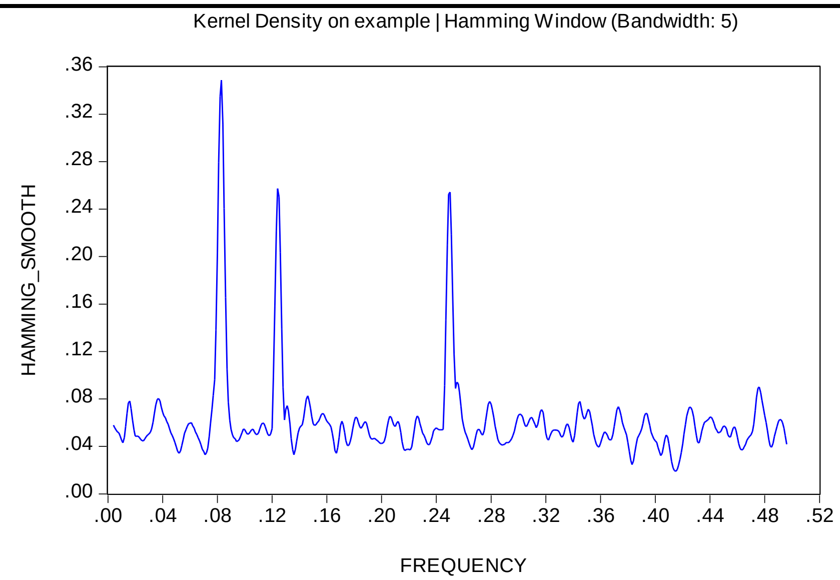
If you click ok, one of the outputs will be a graph object with the frequency f on the abscissa axis and $\sqrt{\widehat{A_i^2} + \widehat{B_i^2}}$ on the ordinate axis. It is important to mention that in calculus of the periodogram the series is transformed to have *zero mean*, as supposed in (1). It can be seen that the frequencies expected to protrude in the spectrum of x actually protruded (Graph 4).

Graph 4 — *Periodogram* —



By default the other outputs will be nine graphs with the smoothed periodogram by their respective window (in graph 5 is showed the smothed periodogram using hamming) and a table called "data" containing the tested angular frequency, the frequency (f), the cycle frequency ($\frac{1}{f}$), the estimated periodogram and the spectral or kernel density data by all windows. If you change the option periodogram smoothing window from "all" to any other window, the output will be the smoothed graph and the table, but the last will only contain the respective kernel density data chosen, also you can change the bandwidth in the edit field. Another useful tool is the decomposition of the periodogram showed in (4) and (5); this can be done by clicking the periodogram components checkbox, and as the periodogram smoothed window, the data table is linked to this option, so the data from A and B will appear in the data table.

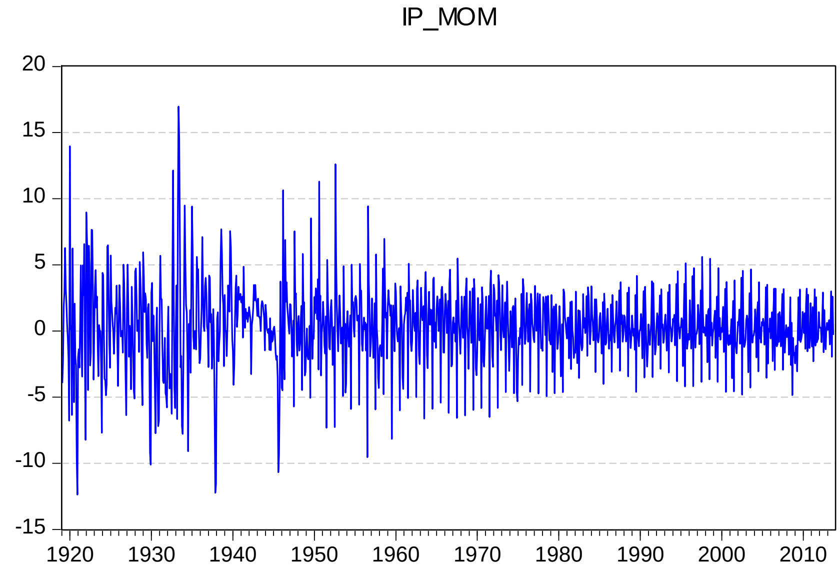
Graph 5 —*Hamming*—



Example 2

This example is focused on real economic data. Specifically the periodogram will be calculated for the industrial production of the United States from 1919 to 2013 with monthly data. This series is showed in graph 6.

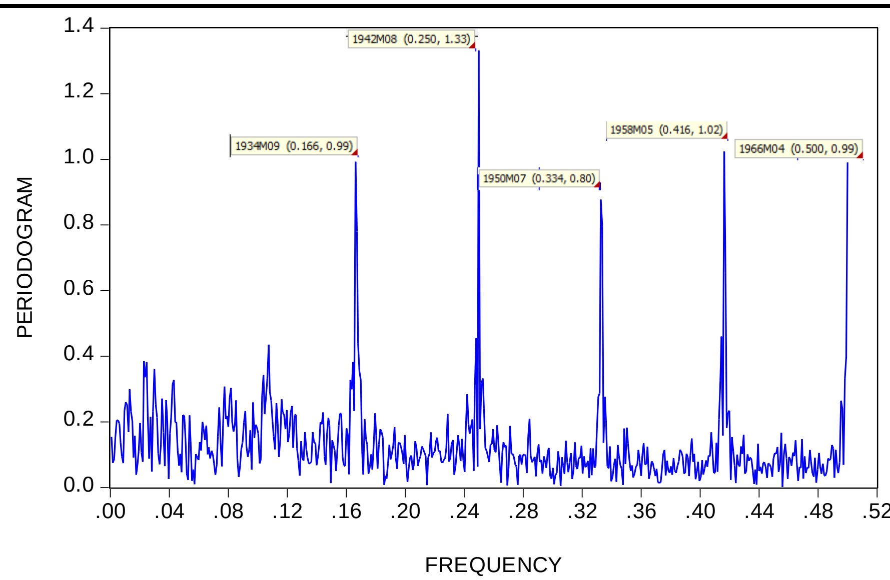
Graph 6 – *Industrial Production %MoM* –



Source: EconStats.

The estimated periodogram for this series is showed in graph 7. The frequencies 0.166, 0.250, 0.334, 0.416 and 0.500 protrude in the spectrum. These frequencies are given in cycles per time unit, then it is inverse $\left(\frac{1}{f}\right)$ is given in time unit per cycle, in this case, months per cycle. Therefore, the given frequencies say that in the industrial production exists a six month, fourth month, three month, 2.4 month and two months cycle.

Graph 7 – *Industrial production periodogram* –



To generate the cycles suggested by the periodogram, there are two ways, the first one is using the structure given in the estimation of the periodogram, that is, the sum of a sine and cosine terms with the same frequency. Therefore, to generate the cycles with these frequencies should be proceed as follows:

*Genr six=@cos(2*3.14159*@trend*0.166)+@sin(2*3.14159*@trend*0.166)*

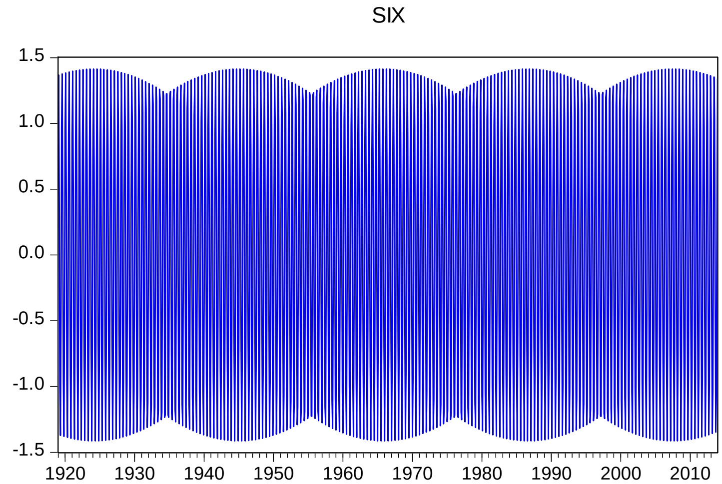
*Genr fourth =@cos(2*3.14159*@trend*0.250)+@sin(2*3.14159*@trend*0.250)*

*Genr three =@cos(2*3.14159*@trend*0.333)+@sin(2*3.14159*@trend*0.333)*

*Genr two_24=@cos(2*3.14159*@trend*0.416)+@sin(2*3.14159*@trend*0.416)*

*Genr two =@cos(2*3.14159*@trend*0.500)+@sin(2*3.14159*@trend*0.500)*

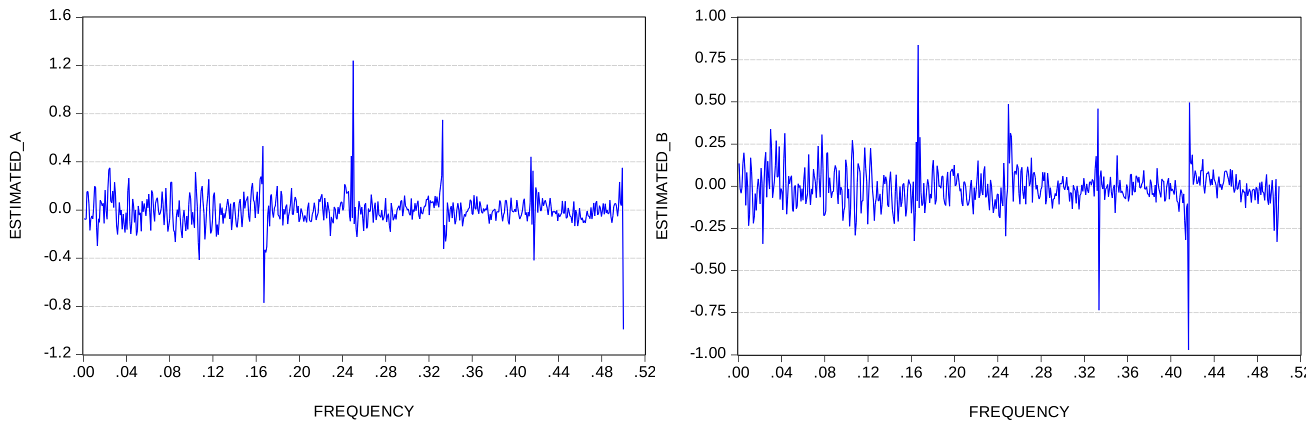
Graph 8 — *Six month cycle* —



The six month cycle it observed in graph 8, their appearance is due to the sum of two signals with close frequencies (in this case equal), this effect it is called *beats*. The second way to generate the cycle is choosing one sine or a cosine function and generate the signal, the choosing criterion could be use the estimates of A and B and see which estimate explains most of the variance of the series[4]. Graph 9 shows the estimates of A and B for the industrial production.

4 Where A represents the covariance with de series and the cosine function, and B with the sine function.

Graph 9 – *A and B estimates* –

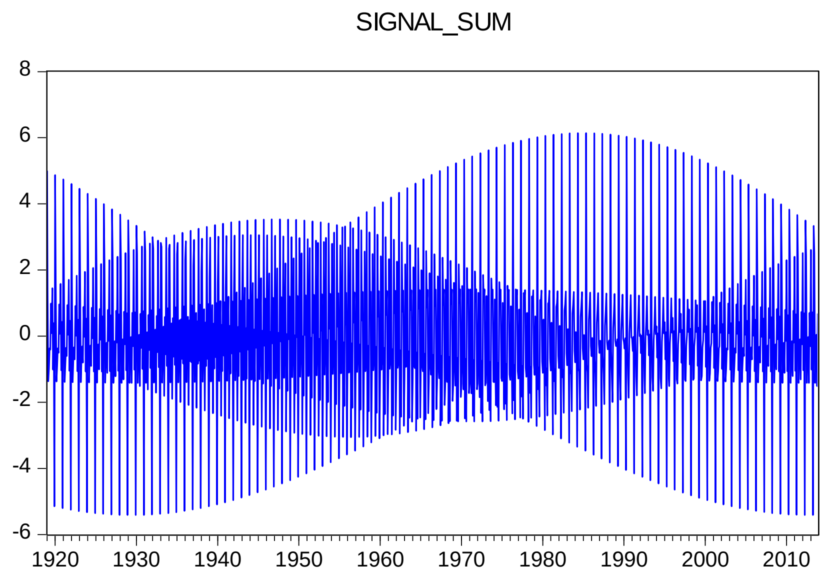


For example, the two month cycle (0.500 frequency) in the periodogram is mostly explained by the A estimates, so it could be an option to generate this cycle only with the cosine function as follows.

*Genr two_a =@cos(2*3.14159*@trend*0.500)*

The selection criteria to generate the cycle should be accompanied by cross-correlation analysis or some way to measure the relation between de series and the cycle, and so the analyst may make a better decision on which way to choose.

Graph 10 – *Signal sum* –



These cycles could be used in forecasting series in the time domain, as the cycle is deterministic and it provides advanced information to the model. In order to obtain parsimony all the cycles could be added to generate only one variable to introduce in the model. In graph 10 it is showed the sum of six, fourth, three, two_24 and two with the sum of sine and cosine terms. This variable could be introduced in a ARMAX(p,q) model as an exogenous variable.

References

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton: Princeton university press.

Montenegro, A. (2009). *Análisis espectral* (No. 007752). UNIVERSIDAD JAVERIANA-BOGOTÁ.

Stoica, P., & Moses, R. L. (2005). *Spectral analysis of signals*. Upper Saddle River, NJ: Pearson/Prentice Hall.