Estimating the spectrum

Now consider quality of $|\hat{X}|^2$ as estimate of f_X.

Have already shown that

$$\mathrm{E}(|\hat{X}|^2) \to f_X(\omega)\,.$$

But variance of this estimate $\hat{f}$ of f does *not* go to 0.

So $\hat{f}$ is not consistent.

Simple case: normal mean 0 process X.

So real and imaginary parts of $\hat{X}$ have normal distributions.

Both have mean 0. Variances are

$$\frac{1}{T}\sum_{r=0}^{T-1}\sum_{s=0}^{T-1}\cos(2\pi\omega r)\cos(2\pi\omega s)C_X(r-s)$$

and

$$\frac{1}{T}\sum_{r=0}^{T-1}\sum_{s=0}^{T-1}\sin(2\pi\omega r)\sin(2\pi\omega s)C_X(r-s)$$

Covariance between real and imaginary parts is

$$\frac{1}{T}\sum_{r=0}^{T-1}\sum_{s=0}^{T-1}\cos(2\pi\omega r)\sin(2\pi\omega s)C_X(r-s)\,.$$

Example: covariance.

Write covariance as

$$\frac{1}{4T}\sum_{r=0}^{T-1}\sum_{s=0}^{T-1} C_X(r-s)\times \\ (e^{2\pi\omega ri}+e^{-2\pi\omega ri})(e^{2\pi\omega si}-e^{-2\pi\omega si})\,.$$

Change variables $u = r - s$ and $v = r + s$ in double sum.

Variable u runs from $-(T-1)$ to $T-1$.

When u is fixed possible values of v run, for $u > 0$ from u to $2(T-1)-u$ by increments of 2.

For $u < 0$ v goes from $-u$ to $2(T-1)+u$ by increments of 2.

For each value of u there are then $T-|u|$ possible values of v.

Covariance becomes

$$\frac{1}{4T}\sum_{u=-(T-1)}^{T-1}\sum_{v=|u|,v\text{ even}}^{2(T-1)-|u|} C_X(u)\times \left\{e^{2\pi\omega vi}-e^{-2\pi\omega vi}+e^{-2\pi\omega ui}-e^{2\pi\omega ui}\right\}.$$

The last two terms, involving u only, are

$$\frac{1}{4T}\sum_{-(T-1)}^{T-1}(T-|u|)C_X(u)(e^{-2\pi\omega ui}-e^{2\pi\omega ui})$$

The terms u and $-u$ cancel each other while the term with $u=0$ is 0 itself so that this term is 0.

Terms involving v simplified using geometric series to do inside sums over v.

Result: coeff of $C(u)$ is bounded (by $4/(1-\cos(2\pi\omega))$ for instance).

Then since

$$\frac{1}{4T}\sum_{u=-(T-1)}^{T-1}|C_X(u)| \to 0$$

covariance between real and imaginary parts of $\hat{X}(\omega)$ converges to 0 as $T \to \infty$.

Mimic calculation of $E\{|\hat{X}(\omega)|^2\}$.

Show variances each converge to $f_X(\omega)/2$.

Hence vector $\sqrt{2/f(\omega)}(Real(\hat{X}(\omega)), Im(\hat{X}(\omega)))$ converges to a bivariate standard normal.

Squared length of this vector then converges in distribution to the squared length of a standard bivariate normal.

This is χ^2_2 or exponential with mean 2.

Summary: $|\hat{X}(\omega)|^2$ converges in distribution to an exponential random variable with mean $f(\omega)$. In particular, $|\hat{X}(\omega)|^2$ is *not* a consistent estimator of $f(\omega)$.

Improved estimates

Two possible approaches: parametric estimation or smoothing.

Here: smoothing.

If $f(\omega)$ is smooth in the neighbourhood of some ω_0 then we can take estimates of $f(\omega)$ at a number of points nearby to ω_0 and average them somehow.

Averaging: reduces variance, introduces bias.

Things being averaged have different expected values.

Simplest smoother is moving average:

$$\hat{f}(k/T) = \frac{1}{2L+1} \sum_{\ell=-L}^{L} |\hat{X}((k+\ell)/T)|^2$$

Fact: quantities being averaged are asymptotically independent.

Estimate has same distribution as an average of $2L+1$ exponentials.

This is chi-squared with $4L+2$ degrees of freedom multiplied by $f(\omega_0)/(4L+2)$.

Can produce consistent estimate by letting L grow slowly with T.

Other weighted averages are possible.

Several implemented in S-Plus function *spectrum*.

Here are some points to note about this estimation problem:

Each estimate $|\widehat{X}(k/T)|^2$ has expected value $f(k/T) + \text{Bias}_T(k/T)$.

Complicated formula for bias can be deduced from algebra above.

Expected value of an estimate of the form

$$\sum_{\ell=-L}^{L} w_\ell |\widehat{X}((k+\ell)/T)|^2$$

is then

$$\sum_{\ell=-L}^{L} w_\ell f((k+\ell)/T) + \sum_{\ell=-L}^{L} w_\ell \text{Bias}_T((k+\ell)/T)$$

If f is roughly linear around k/T: first term will be quite close to $f(k/T)$ when the weights make the estimate an average, that is, they sum to 1.

Approximation will be poor in neighbourhood of any peak in spectrum.

Peaks will be flattened by averaging.

Second term in expectation has no particular reason to average out to 0.

Increasing L without dealing with this bias is eventually fruitless: bias becomes dominant component in error.

Common tactic: *tapering*

Compute

$$\hat{X}^*(\omega) = \sum h\left(\frac{t}{T}\right) X_t \exp(2\pi\omega t i)$$

and use as a periodogram

$$\frac{|\hat{X}^*(\omega)|^2}{\sum h^2\left(\frac{t}{T}\right)}$$

where tapering function h typically decreases to 0 at 0 and at 1.

Motivation for tapering:

Begin with Fourier expansion of a function f:

$$f(\omega) = \sum_{h=-\infty}^{\infty} a_h \exp(2\pi\omega i h)$$

In practice approximate with partial sum

$$f_n(\omega) = \sum_{h=-n}^{n} a_h \exp(2\pi\omega i h)$$

Recall formula for coefficients:

$$a_h = \int_{-0.5}^{0.5} f(\omega) \exp(-2\pi\omega i h) d\omega$$

Substitute in expansion of f_n to find

$$f_n(\omega_0) = \int_{-0.5}^{0.5} f(\omega) \sum_{h=-n}^{n} \exp\{2\pi i(\omega_0 - \omega)h\} d\omega$$

Sum inside integral has form

$$\sum_{-n}^{n} \rho^k = \frac{\rho^{-n} - \rho^{n+1}}{1-\rho}$$

where $\rho = \exp\{2\pi i(\omega_0 - \omega)\}$.

Put $\theta = 2\pi(\omega_0 - \omega)$ then sum is

$$\frac{e^{(n+1)\theta i} - e^{-n\theta i}}{1 - \exp(\theta i)}$$

Multiply top and bottom by $\exp(i\theta/2)$ to get

$$D_n(\omega_0 - \omega) \equiv \frac{\sin\{2\pi(n+1/2)(\omega - \omega_0)\}}{\sin\{2\pi(\omega - \omega_0)/2\}}$$

(Notice D_n is even.) Thus:

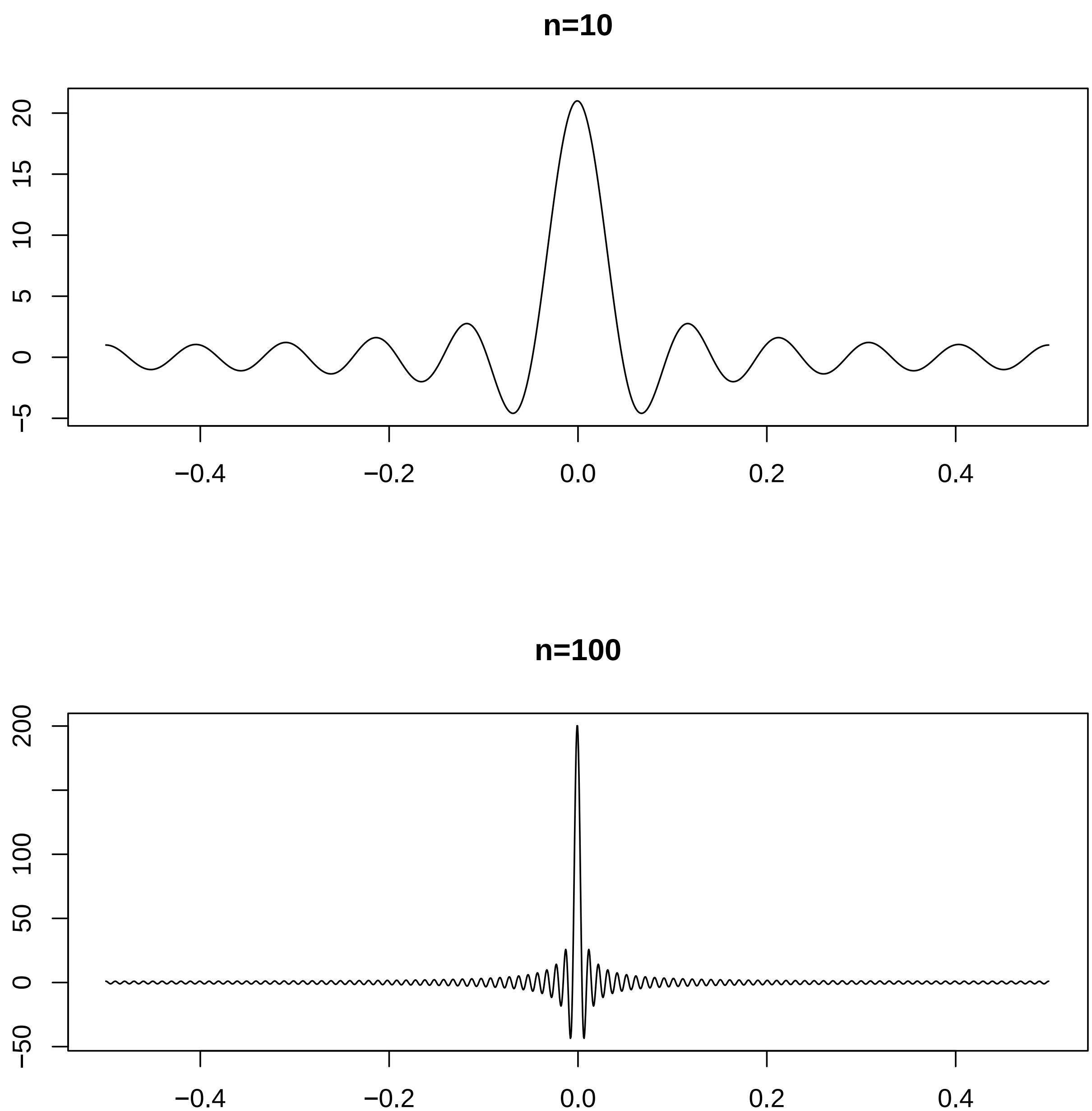
$$f_n(\omega_0) = \int_{-0.5}^{0.5} f(\omega) D_n(\omega_0 - \omega) d\omega$$

Note: D_n is periodic with period 1 because it is linear combination of $\exp(2\pi i k \omega)$.

Also

$$\int_{-0.5}^{0.5} D_n(\omega_0 - \omega) d\omega = 1$$

Plots of D_n for $n = 10, 100$.



Things to notice:

Since $\int D_n = 1$ we see $f_n(\omega_0)$ is weighted average of $f(\omega)$.

Major spike in center so most weight near ω_0.

Some negative weights. (Probably undesirable.)

Secondary peaks in D_n at places where $\sin\{2(n+1/2)\pi\theta)\} = 1$.

If $\pm\alpha_0 = \pm(4n+2)^{-1}$ closest such to 0 and f has spike at ω_0 then f_n has smaller spike at $\omega_0 \pm \alpha_0$. Spike is artificial.

Partial sum of Fourier series optimal in sense of integrated squared error.

Other suggestions made to avoid spikes, however.

Idea: if $a_n \to a_\infty$ then

$$\frac{a_0 + \cdots + a_n}{n+1} \to a_\infty$$

If a_n values oscillate above and below a_∞ averages might converge faster.

Apply to f_n:

$$\tilde{f}_n = \frac{f_0 + \cdots + f_n}{n+1}$$

Use formula for f_k to get formula for $\tilde{f}_n$:

$$\tilde{f}_n = (n+1)^{-1} \sum_{h=-n}^{n} C(h)(1-|h|/n)\exp(2\pi\omega i h)$$

Notice result is less weight on $C(h)$ for large h.

Again: use Fourier formula for C_h to write

$$\tilde{f}_n(\omega) = \int_{-0.5}^{0.5} f(\omega) K_n(\omega_0 - \omega) d\omega$$

where

$$K_n(\alpha) = \sum_{h=-n}^{n} (1 - |h|/n) \exp(2\pi i h \alpha)$$

Can do this some analytically; see Brillinger's book.

Other suggestions amount to

$$\tilde{f}_n = (n+1)^{-1} \sum_{h=-n}^{n} C(h) g(|h|/n) \exp(2\pi \omega i h)$$

for a function g with $g(0) = 1$ and g continuous at 0.

Call g a taper or data window or convergence factor. Get other weight functions like K_n.

Source of jargon for data-window, taper?

Note formula for $|\hat{X}(\omega)|^2$.

$$\begin{aligned}
|\hat{X}(\omega)|^2 &= \hat{X}(\omega)\overline{\hat{X}(\omega)} \\
&= \frac{1}{T}\sum_{s=0}^{T-1}\sum_{t=0}^{T-1} X_s X_t \exp\{2\pi i\omega(s-t)\} \\
&= \frac{1}{T}\sum_{h=-(T-1)}^{T-1}\sum_{t=0}^{T-1-|h|} X_t X_{t+|h|} \exp(\pi i\omega h) \\
&= \sum_{h=-(T-1)}^{T-1} \hat{C}(h) \exp(\pi i\omega h)
\end{aligned}$$

Notice that DFT of estimated autocovariance is modulus squared of DFT of data.

Here

$$\hat{C}(h) = \frac{1}{T}\sum_{t=0}^{T-1-|h|} X_t X_{t+|h|}$$

which is the unbiased estimate multiplied by the convergence factor $1 - |h|/T$.

Idea?

For a reasonable stationary series $C(h)$ goes to 0 as $h \to \infty$ quite quickly.

So: for h so large that SE large compared to likely values of $C(h)$ use **shrinkage**.

Multiply $\widehat{C}(h)$ by weight shrinking towards 0.

I.e. use convergence factor in DFT of autocovariance.

Ideal time to smooth periodogram is when the spectrum is flat, that is, when the series is white noise.

If $\mathcal{A}$ is a filter such that $Y = \mathcal{A}(X)$ is nearly white noise then we could

1. Transform X to Y.

2. Compute the periodogram of Y.

3. Smooth this periodogram fairly heavily, because there should be no significant peaks in f_Y. Call the resulting estimate $\widehat{f}_Y$.

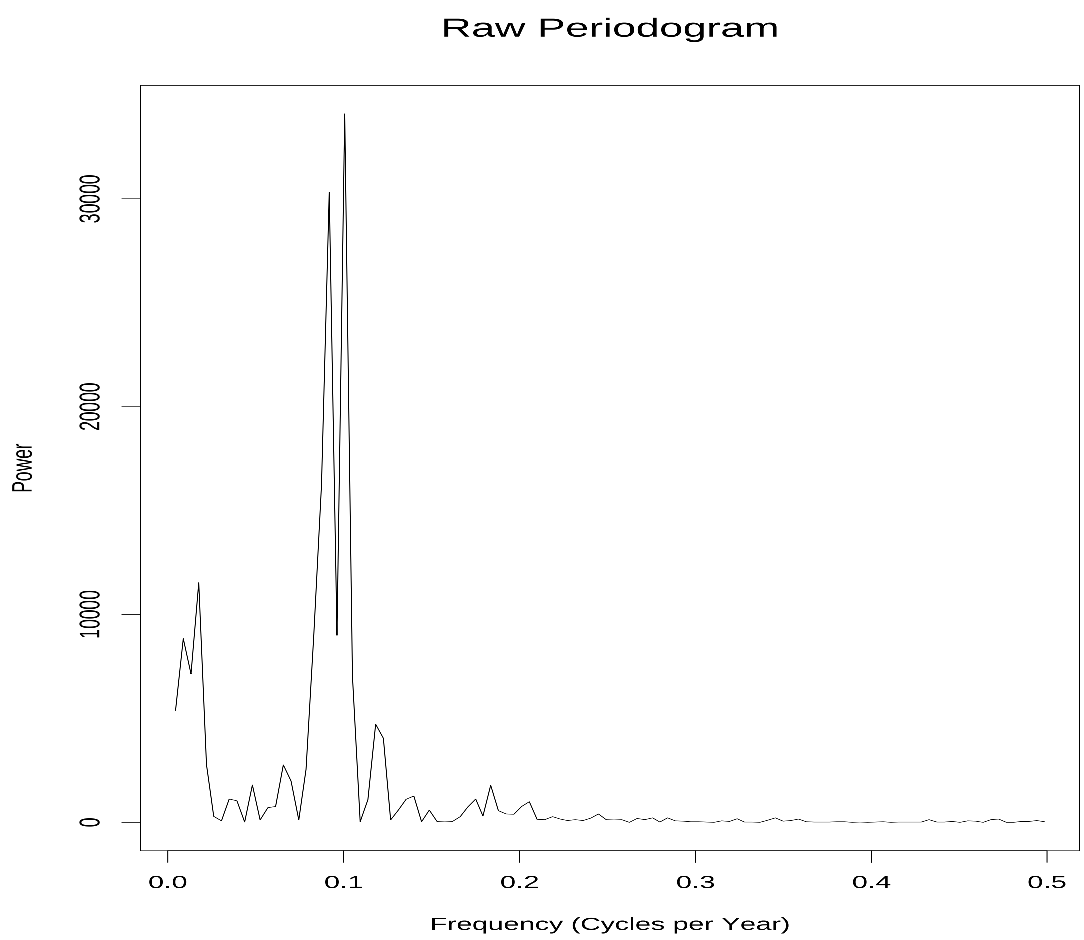
4. Estimate f_X by

$$\widehat{f}_X(\omega) = \frac{\widehat{f}_Y(\omega)}{|A(\omega)|^2}$$

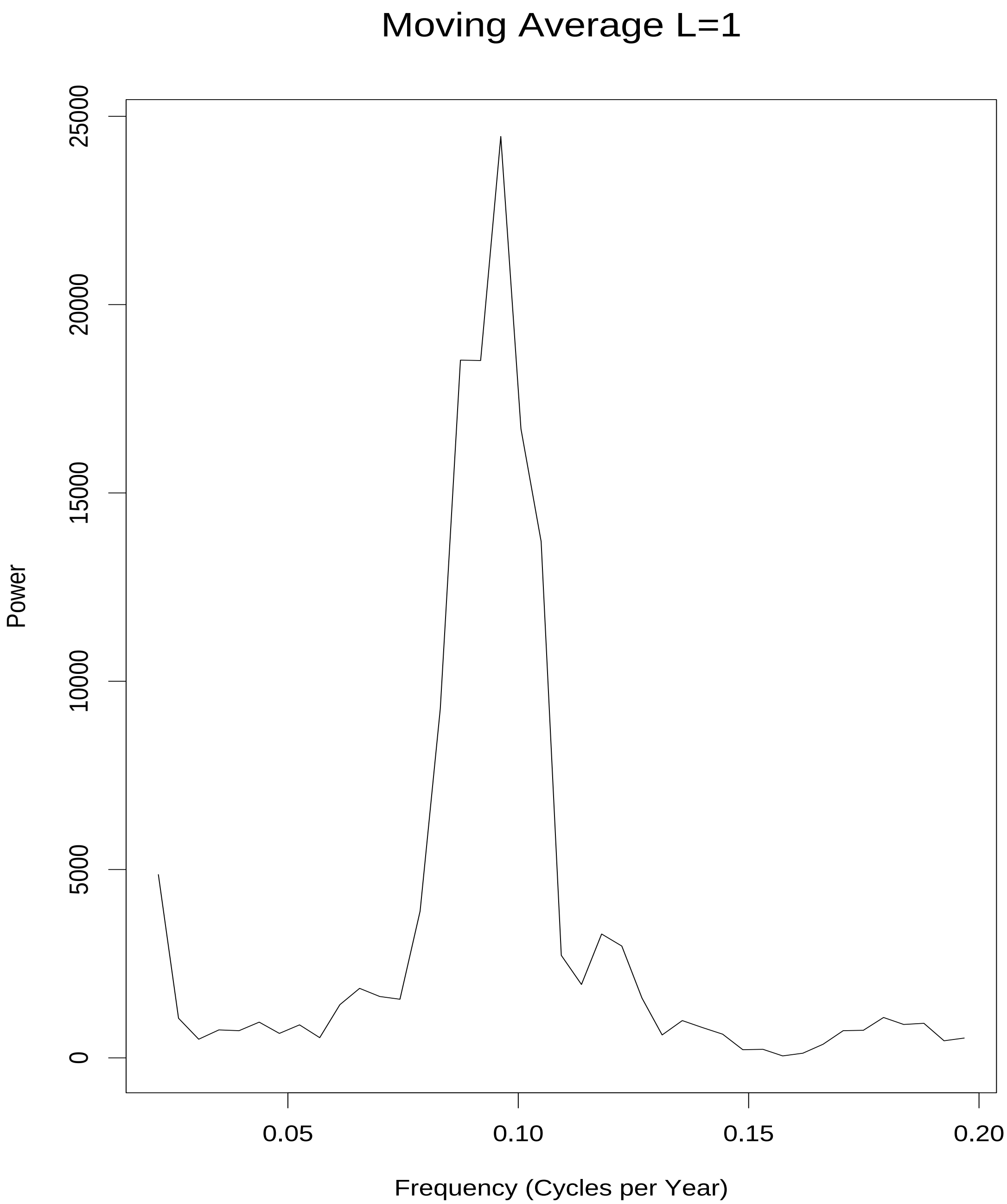
   where A is the frequency response function of the filter $\mathcal{A}$.

Several spectral estimates for sunspots series:

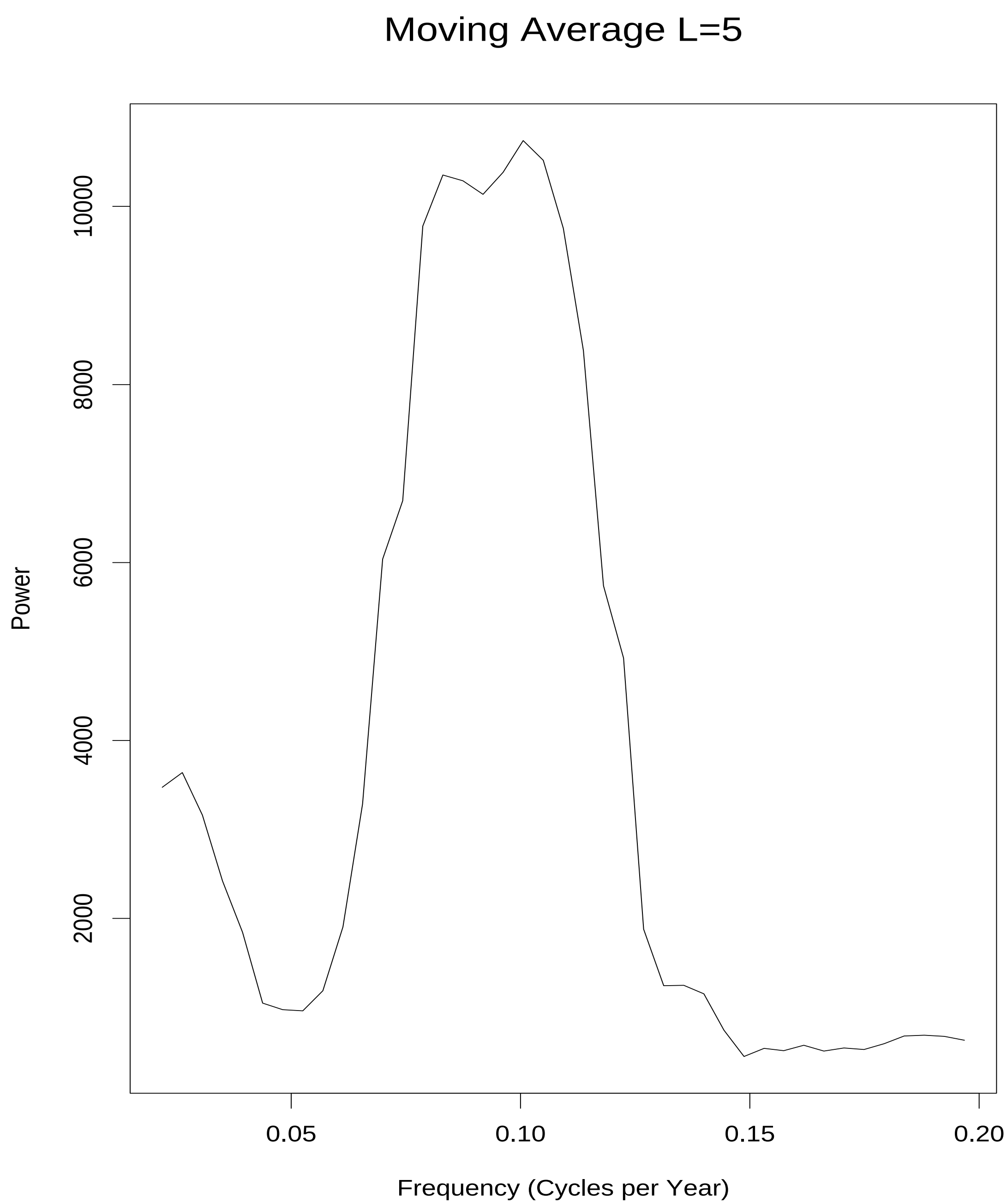
Raw periodogram. Are there two peaks near a period of 10 years? Is there a peak near 40 years?



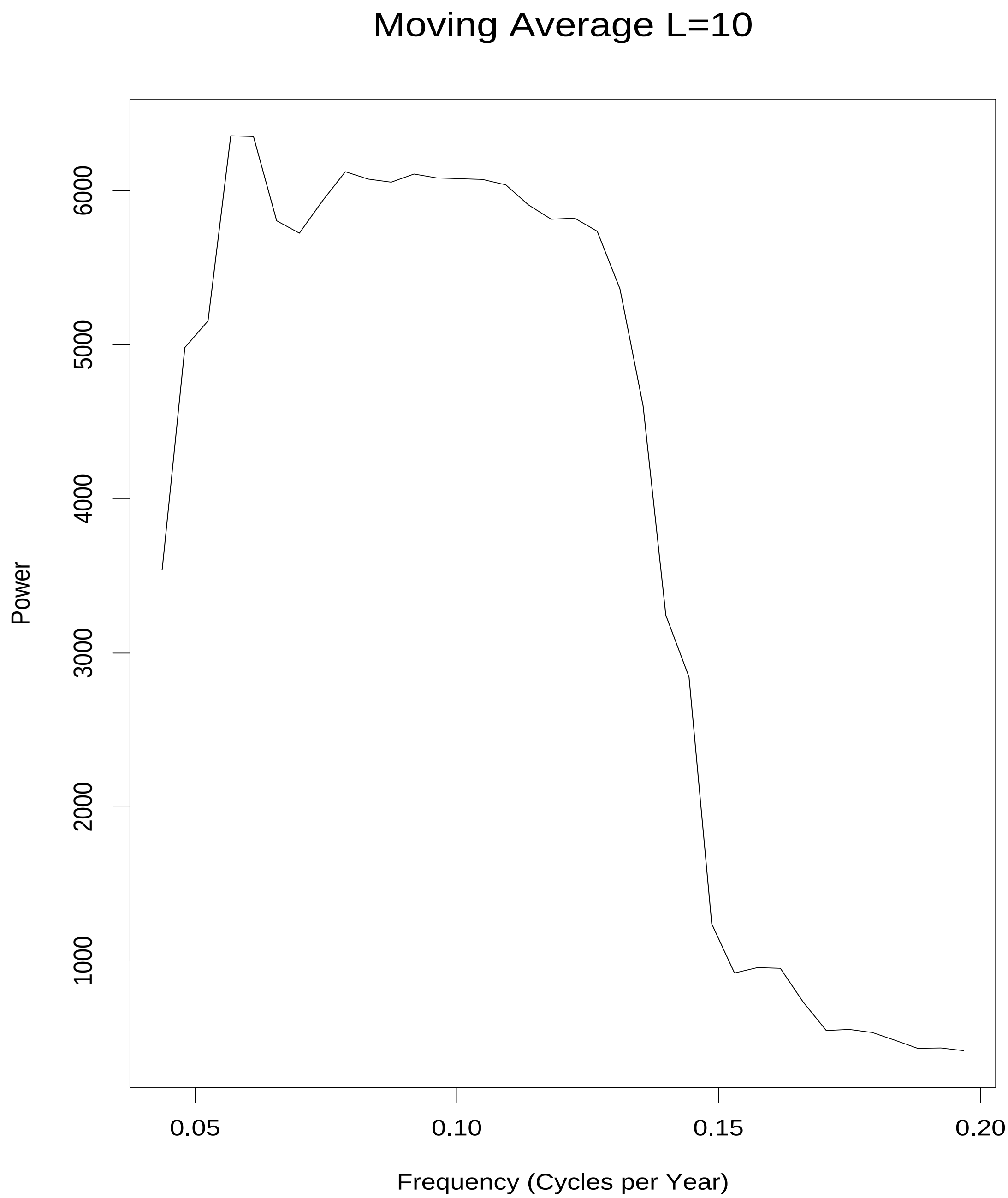
Running means with $L = 1$.



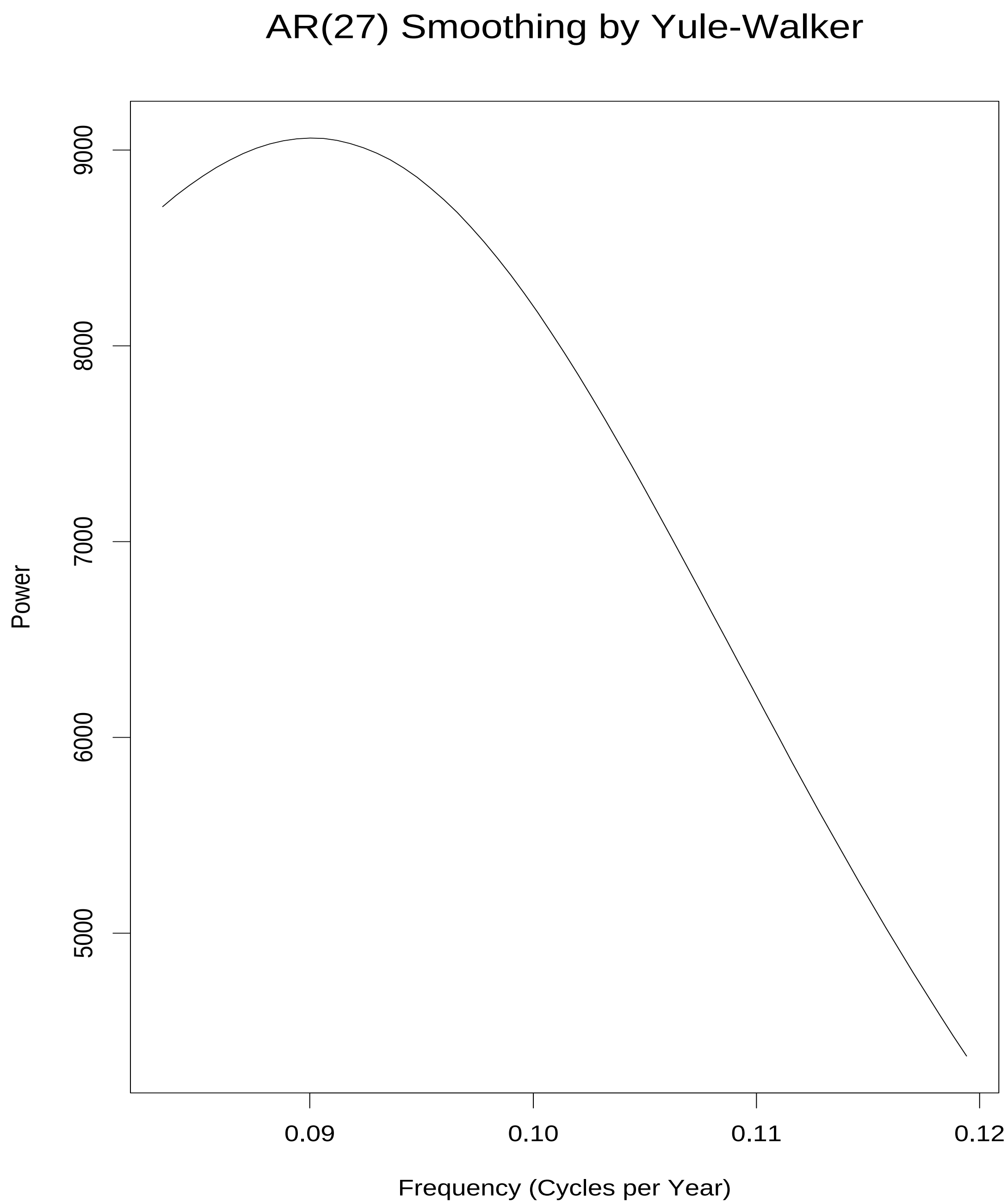
Running means $L = 5$.



Running means $L = 10$.



Prewhitening by AR(27) model:



Prewhitening by a high order AR(1000).