Stat 565

Estimating The Spectrum

Feb 23rd 2016

Charlotte Wickham

stat565.cwick.co.nz

Today

- Finish up spectrum
- Estimating the spectrum: the "periodogram"
- Some periodogram analysis examples
- Project time @11:05am

Review

Frequency domain analysis:

We consider time series as the superposition of periodic (cos-sin pairs) functions at different frequencies.

The spectral density function describes how much variance each frequency contributes to the variance of our process.

Spectrum and autocovariance

$$f(\omega) = \frac{1}{\pi}\left[\gamma(0) + 2\sum_{k=1}^{\infty}\gamma(k)\cos\omega k\right]$$

$$\gamma(k) = \int_0^{\pi}\cos(\omega k)\, f(\omega) d\omega$$

The autocovariance function and the spectral density both contain the same amount of information.

Derive spectrum for white noise

Derive spectrum for MA(1)

HW #7

Spectrum for AR(1)

Fourier expansion

For a time series of length N (N even),the finite Fourier series expansion is,

$$
\begin{aligned}
x_t =& a_0 \\
&+ \sum_{p=1}^{(N/2)-1} \left[a_p \cos(2\pi pt/N) + b_p \sin(2\pi pt/N)\right] \\
&+ a_{N/2} cos(\pi t)
\end{aligned}
$$

t = 1, ... , N

$$
\begin{aligned}
a_0 &= \bar{x} \\
a_{N/2} &= \sum (-1)^t x_t / N \\
a_p &= 2\left[\sum x_t cos(2\pi pt/N)\right]/N \\
b_p &= 2\left[\sum x_t sin(2\pi pt/N)\right]/N
\end{aligned}
$$

equivalent to regression of x_t on the cosine-sine pairs at frequencies 2πp/N

Parseval's Theorem

$$1/N \sum_{t=1}^{N} (x_t - \bar{x})^2 = \sum_{p=1}^{(N/2)-1} R_p^2/2 + N a_{N/2}^2$$

$$R_p = \sqrt{a_p^2 + b_p^2}$$

$$\omega_p = 2\pi p/N$$

A plot of $R^2_p/2$ against ω_p is called a line spectrum. Generally we actually plot

$$I(\omega_p) = N R_p^2 / 4\pi$$

We call $I(\omega_p)$ the periodogram.

The periodogram is an estimate of the spectrum

c_k = sample auto covariance at lag k

Can show:

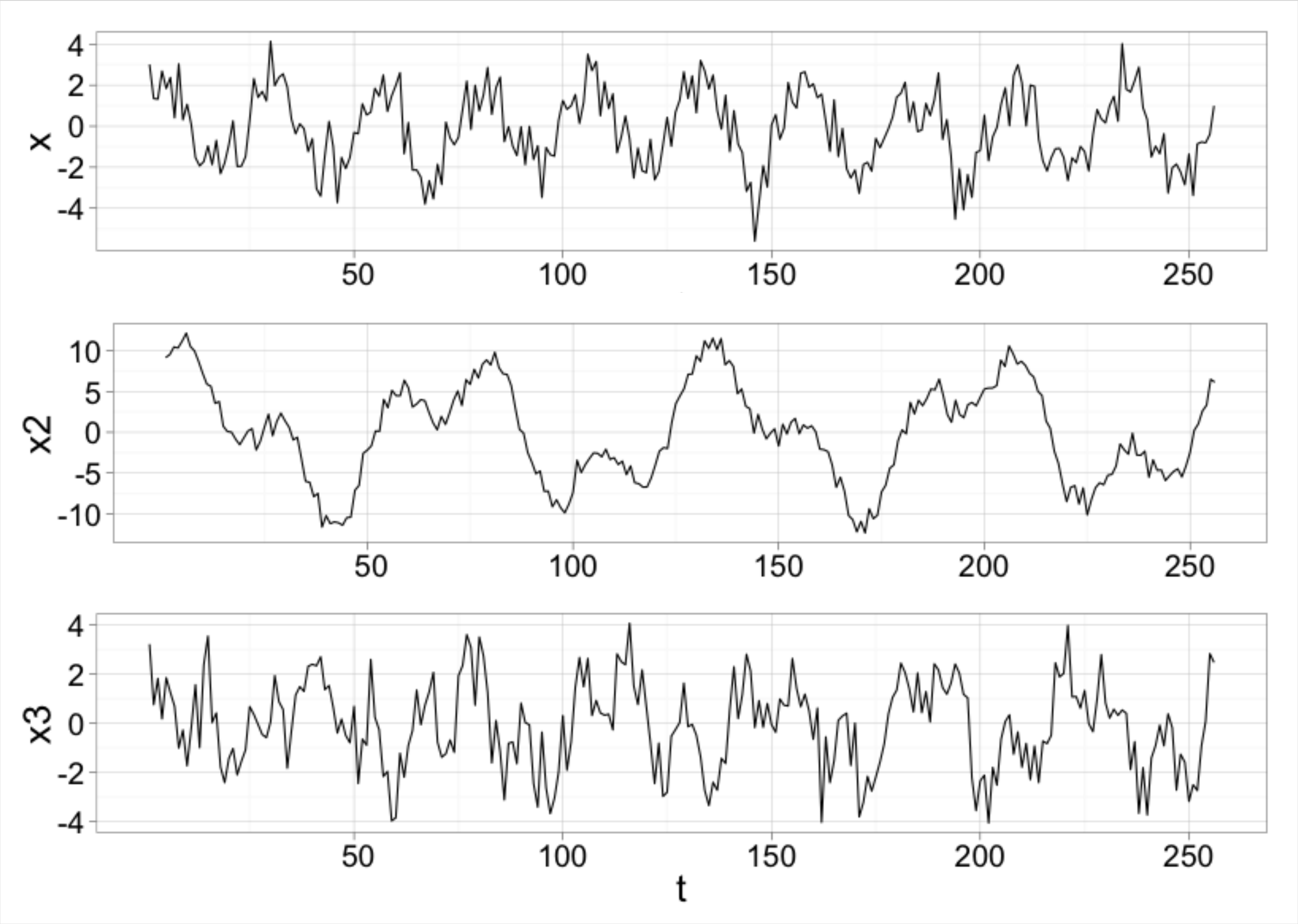
$$I(\omega_p) = \frac{1}{\pi}\left(c_0 + 2\sum_{k=1}^{N-1} c_k \cos\omega_p k\right)$$

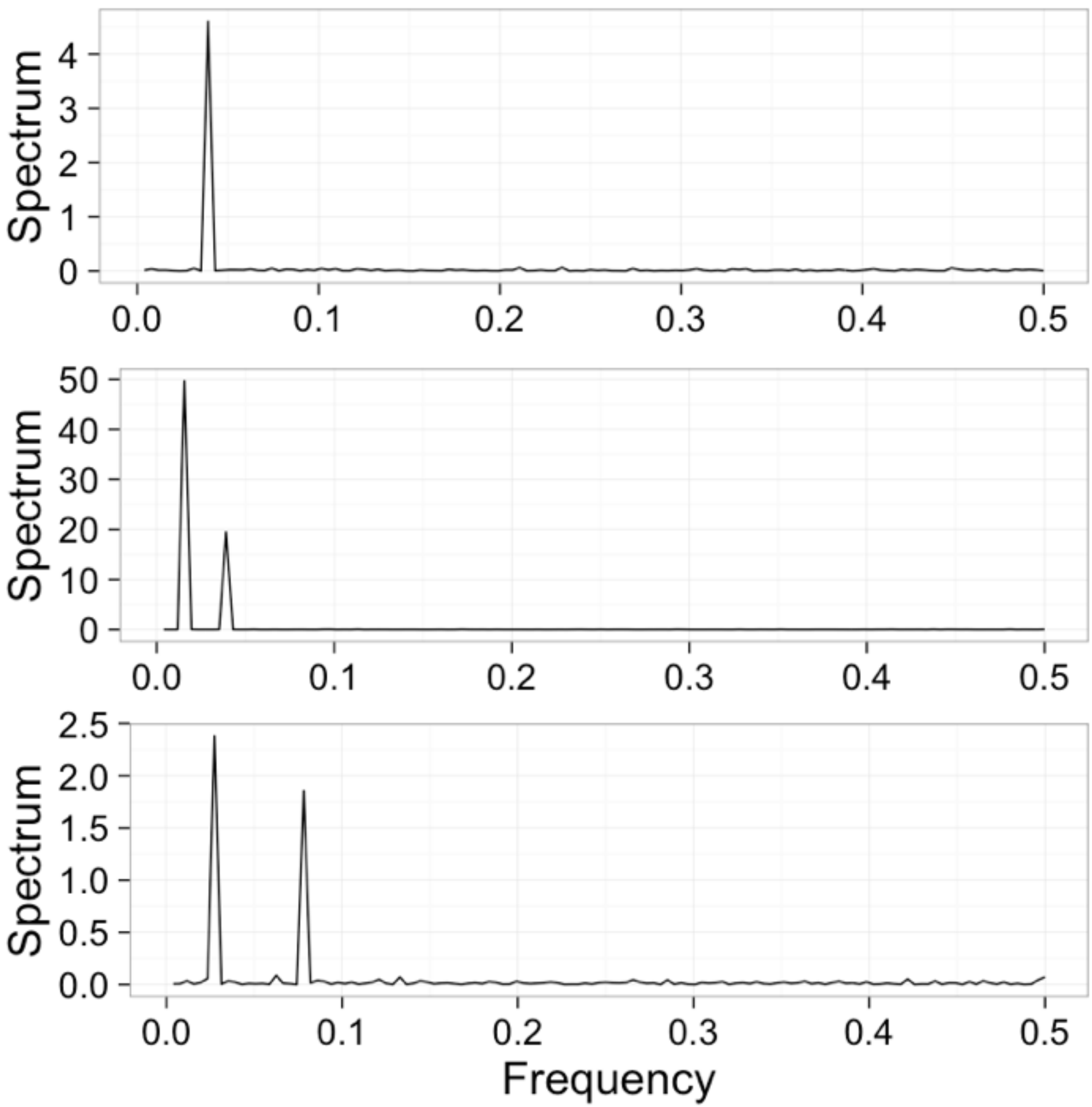
Compare to:

$$f(\omega) = \frac{1}{\pi}\left[\gamma(0) + 2\sum_{k=1}^{\infty}\gamma(k)\cos\omega k\right]$$

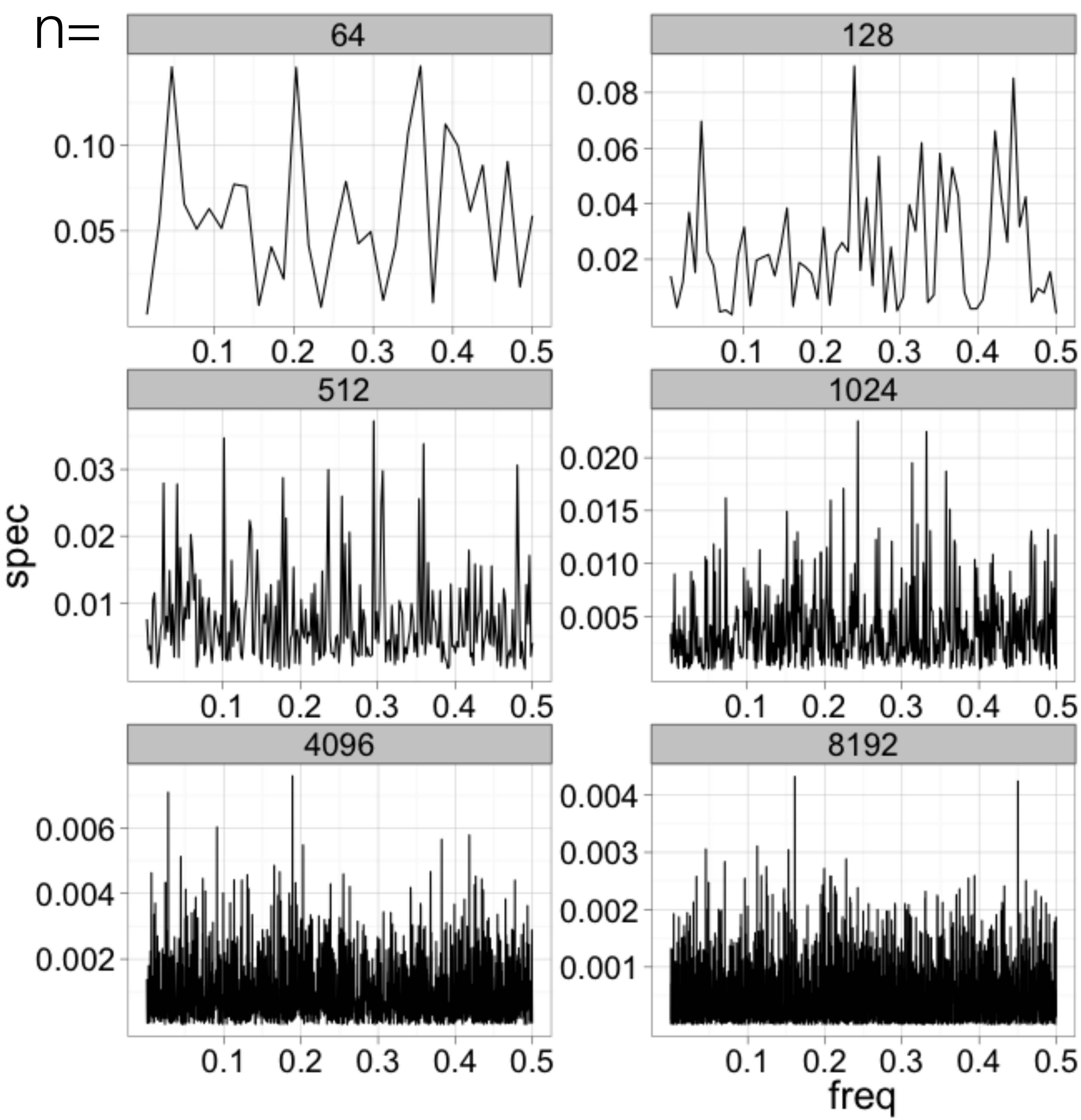
The periodogram can be thought of as the "sample spectrum"

Can you guess the dominant frequencies in these (simulated) series?




Spectrum
4
3
2
1
0
0.0
0.1
0.2
0.3
0.4
0.5
Spectrum
50
40
30
20
10
0
0.0
0.1
0.2
0.3
0.4
0.5
Spectrum
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.1
0.2
0.3
0.4
0.5
Frequency

Periodogram is asymptotically unbiased but not consistent!



spectrum for white noise

Ideally, with a very long series we would estimate a flat spectrum.

Properties of $I(\omega_p)$

$$\frac{2I(\omega_p)}{f(\omega_p)} \rightarrow_D \chi_2^2$$

The periodogram is an asymptotically unbiased estimate of the spectrum.

The periodogram is not a consistent estimator of the spectrum.

$I(\omega_i)$ $I(\omega_j)$ are asymptotically independent, for ω_i & ω_j

Smoothed periodogram

Smoothing allows us to make the estimator consistent but introduces bias.

The simplest case is simply to average neighboring values,

$$\hat{f}(\omega) = \frac{1}{m} \sum_j I(\omega_j)$$

where ω_j are m consecutive Fourier frequencies centered around ω.

In practice, we use a weighted average, giving more weight to frequencies in the middle of the band.

$$\hat{f}(\omega) = \sum_{k=-m}^{m} h_k I(\omega_p + 2\pi k/N)$$

where $\sum_{k=-m}^{m} h_k = 1$ h_k = the kernel

$$L_h = \left(\sum_{k=-m}^{m} h_k^2 \right)^{-1}$$ bandwidth, $B = L_h/n$

Smoothing is subjective

Smoothing reduces variance, but it introduces bias.

We want to reduce variance, without introducing too much bias.

How much smoothing is subjective, and it's worth playing with.

In R: spectrum

By default:

Removes a linear trend

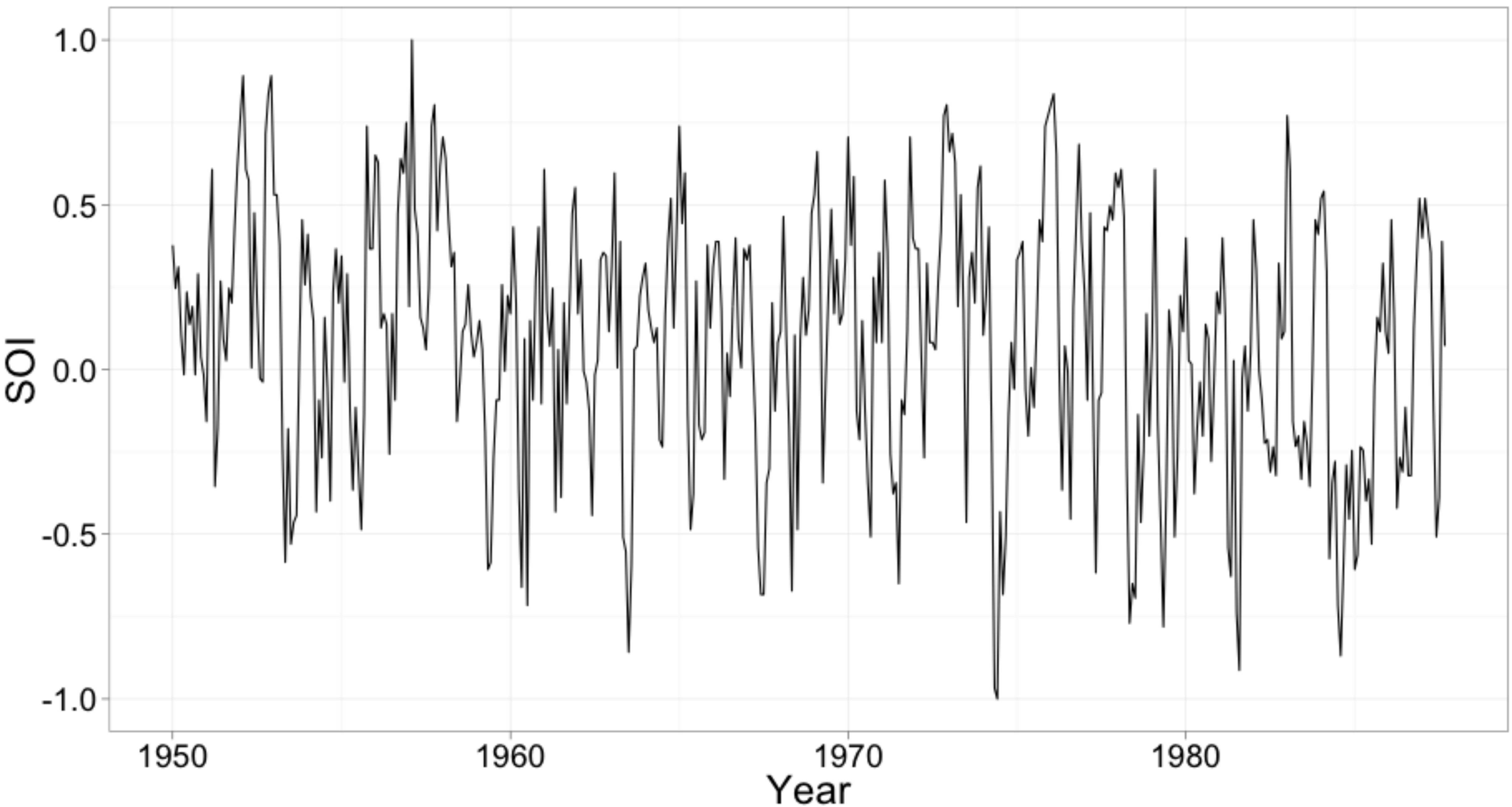
Doesn't smooth

But if you specify a `span` uses a “Modified Daniell” kernel

Plots on a log scale

Tapers (10% of data) when the true frequency occurs between Fourier frequencies, it's power will leak into Fourier frequencies around it, tapering attempts to reduce this (see C&C 14.5 for the best discussion)

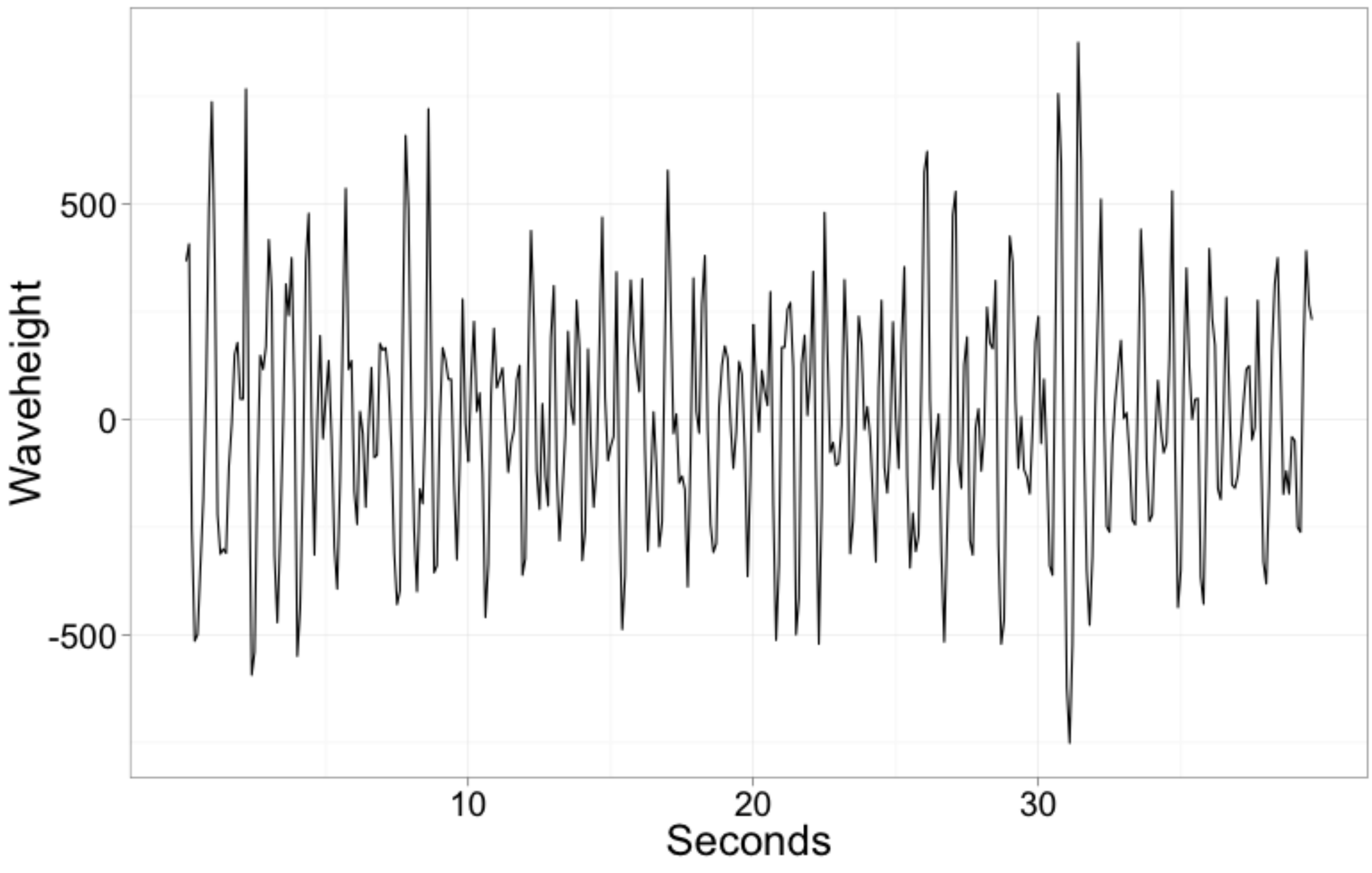
Southern Oscillation Index



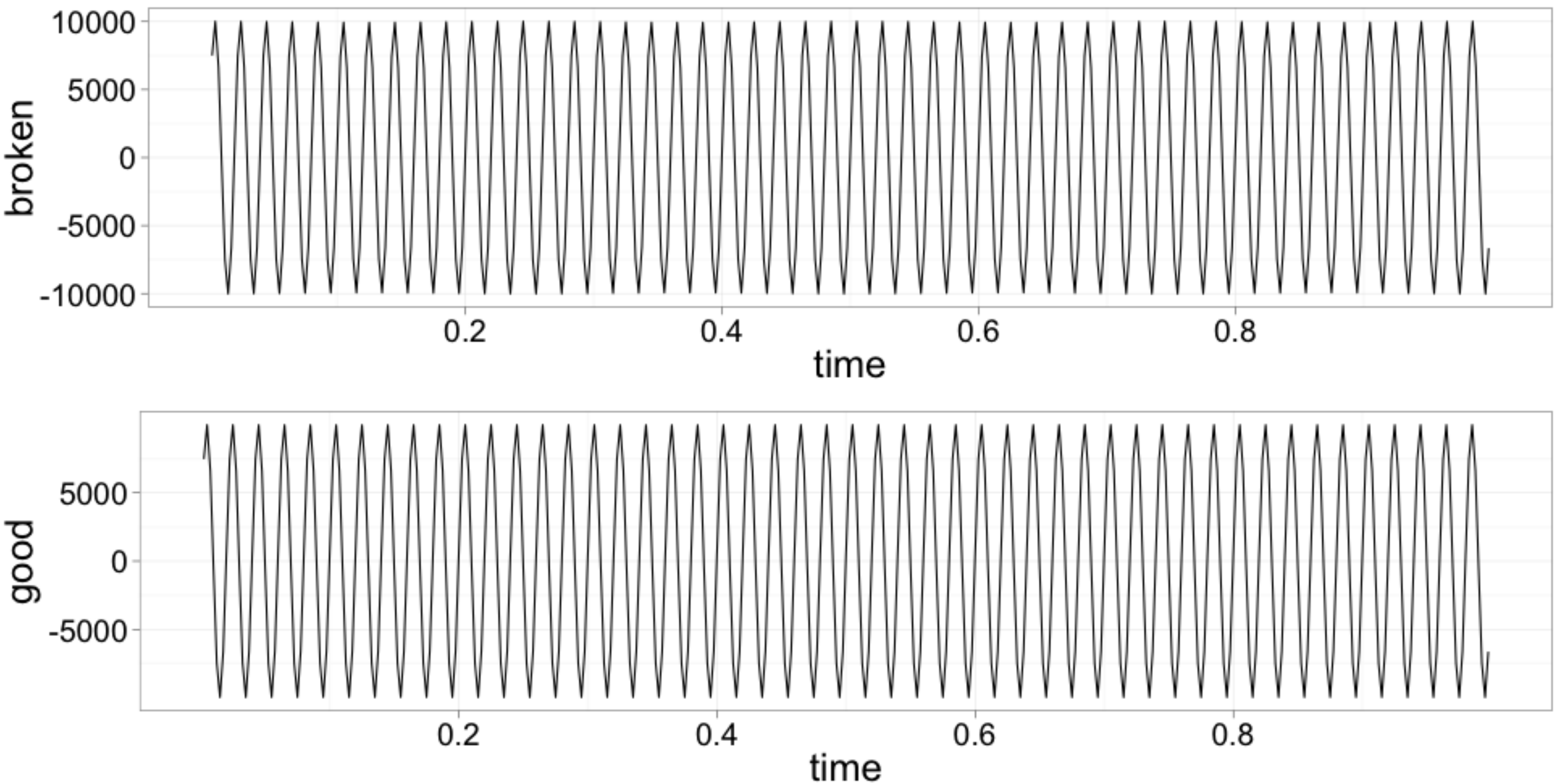
SOI is the pressure difference between Tahiti and Darwin, it should capture El Niño.

```r
spectrum(soi)

# no log scale, no taper, remove mean not trend
spectrum(soi, log = "no", taper = 0, demean = TRUE,
detrend = FALSE)

# averaged periodogram (average over 2*4 + 1 = 9
values)
spectrum(soi, spans = 4, log = "no", taper = 0)

# if you use the log scale you get a confidence
band estimate
spectrum(soi, spans = 4, taper = 0)

# a kernel with more weight in the middle
spectrum(soi, spans = c(3, 3), taper = 0)
```

Height of wave in wave tank



Good and broken motors



A different approach to estimating the spectrum

Fit a high order ARMA(p, q) process and use the relationship between the auto covariance function and spectrum.

See `spec.ar`