Example 1.12, Signal in noise 2.7, using regression to discover a signal in noise 2.8 using the periodogram to discover a signal in noise

Many realistic models for generating time series assume an underlying signal with some consistent periodic variation, contaminated by adding a random noise.

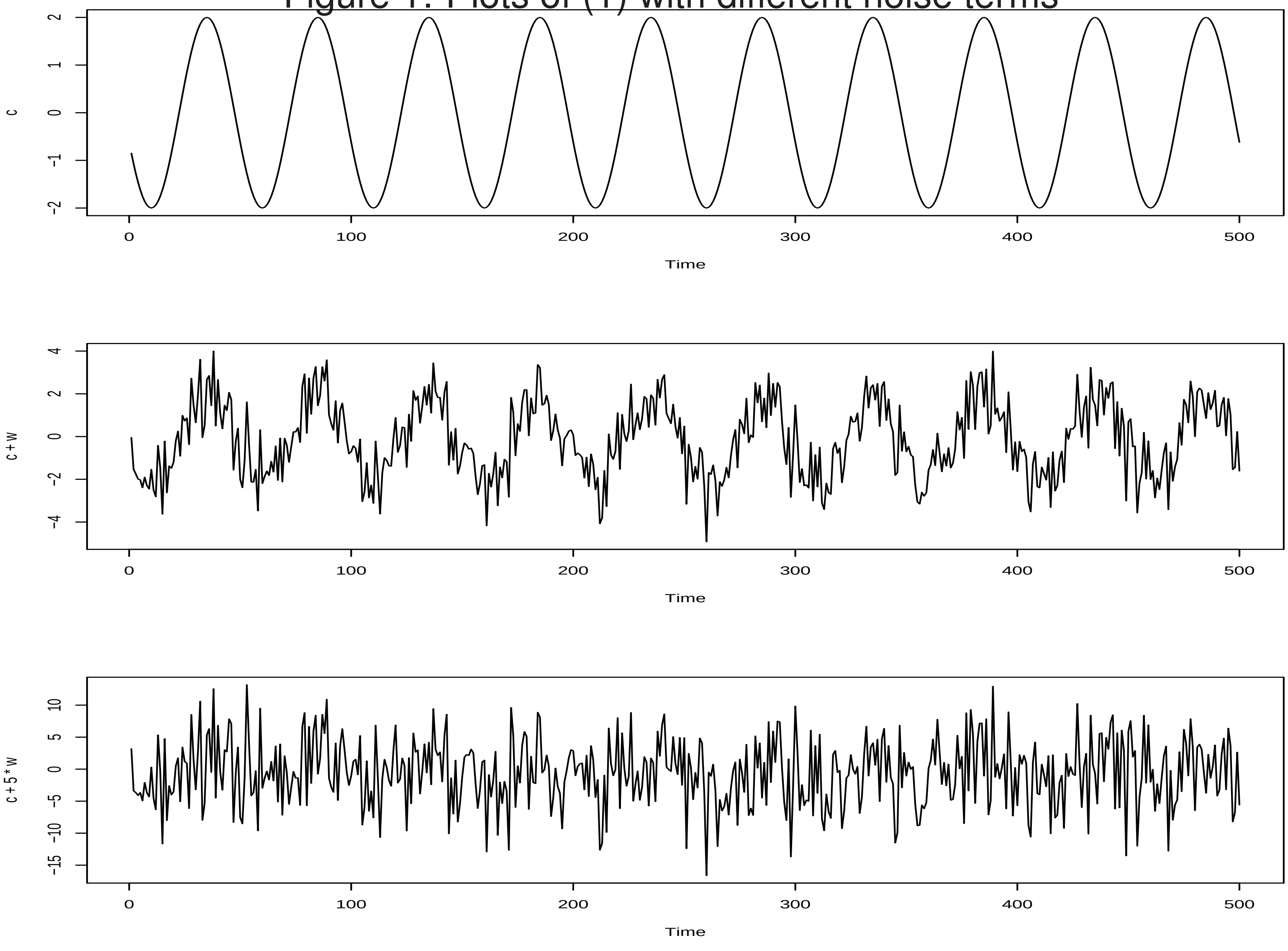
Consider the model

$$x_t = 2cos(2\pi t/50 + .6\pi) + w_t \tag{1}$$

for $t = 1, 2, \cdots, 500$, where the first term is regarded as the signal. shown in the upper panel of Figure 1. An additive noise term was taken to be white noise with $\sigma_w = 1$ (middle panel) and

$\sigma_w = 5$ (bottom panel), drawn from a normal distribution.

Figure 1: Plots of (1) with different noise terms



- The ratio of the amplitude of the signal to σ_w is sometimes called the signal-to-noise ratio (SNR)
- The larger the SNR, the easier it is to detect the signal
- The signal is easily discernible in the middle panel of Figure 1, whereas the signal is obscured in the bottom panel
- Typically, we will not observe the signal, but the signal obscured by noise

Assume that $w = 1/50$ is known, but A and ϕ are unknown parameters. Write (1) in another way

$$x_t = \beta_1 cos(2\pi t/50) + \beta_2 sin(2\pi t/50) + w_t$$

where $\beta_1 = Acos(\phi)$ and $\beta_2 = -Asin(\phi)$.

Using linear regression on the generated data,

- the fitted model is

$$\hat{x}_t = -.84_{(.32)} cos(2\pi t/50) - 1.99_{(.32)} sin(2\pi t/50)$$

with $\hat{\sigma}_w = 5.08$, where the values in parentheses are the standard errors.

- The actual values of the coefficients for this example are $\beta_1 = 2cos(.6\pi) = -.62$ and $\beta_2 = -2sin(.6\pi) = -1.90$
- The parameter estimates are significant and close to the actual values. We are able to detect the signal in the noise using regression.

Using the periodogram to discover a signal in noise

Consider w a parameter,

- Fit model (1) using nonlinear regression with w as a parameter.
- Try various values of w in a systematic way

$$\hat{\beta}_1 = \frac{\sum_1^n x_t cos(2\pi t/50)}{\sum_1^n cos^2(2\pi t/50)} = \frac{2}{n}\sum_{t=1}^{n} x_t cos(2\pi t/50)$$

$$\hat{\beta}_2 = \frac{\sum_1^n x_t sin(2\pi t/50)}{\sum_1^n sin^2(2\pi t/50)} = \frac{2}{n}\sum_{t=1}^{n} x_t sin(2\pi t/50)$$

- Suggests looking at all possible regression parameter estimates

$$\hat{\beta}_1(j/n) = \frac{2}{n}\sum_{t=1}^{n} x_t cos(2\pi tj/n);$$

$$\hat{\beta}_2(j/n) = \frac{2}{n}\sum_{t=1}^{n} x_t sin(2\pi tj/n),$$

where, $n = 500$ and $j = 1, \cdots, n/2 - 1$ and inspecting the results for large values.

- When $j = 0 \quad \text{or} \quad n/2, \hat{\beta}_1(0) = 2n^{-1}\sum_{t=1}^{n} x_t, \hat{\beta}_1(\frac{n}{2}/n) = 2n^{-1}\sum_{t=1}^{n}(-1)^t x_t$ and $\hat{\beta}_2(0) = 0, \hat{\beta}_1(\frac{n}{2}/n) = 0$

- We have regressed a series, x_t of length n using n regression parameters, so that we have a perfect fit.
- $\hat{\beta}_1(j/n)$ and $\hat{\beta}_2(j/n)$ for each j are essentially measure the correlation of the data with a sinusoid oscillating at j cycles in n time points.
- An appropriate measure of the presence of a frequency of oscillation of j cycles in n time points in the data would be

$$P(j/n) = \hat{\beta}_1^2(j/n) + \hat{\beta}_2^2(j/n),$$

which is basically a measure of squared correlation. $P(j/n)$ is called periodogram or scaled periodogram.

Figure 2: The scaled periodogram of the 500 observations generated by (1).