Least Squares Fitting of Sinusoidal Models

Suppose we suspect that our data is dominated by a periodic component at frequency $\omega$. Assume for now that $\omega$ is known. Then it is appropriate to fit the model

$$x_t = \mu + A \cos \omega t + B \sin \omega t + \varepsilon_t,$$

(1)

where the $\varepsilon_t$ are random error terms, typically assumed to be uncorrelated with zero mean and equal variances. The exact least squares estimates of the parameters $\mu, A, B$ are solutions to the normal equations

$$\sum (x_t - \mu - A \cos \omega t - B \sin \omega t) = 0,$$

$$\sum \cos \omega (x_t - \mu - A \cos \omega t - B \sin \omega t) = 0,$$

$$\sum \sin \omega (x_t - \mu - A \cos \omega t - B \sin \omega t) = 0.$$

The terms in these equations which do not depend on $x_t$ may be evaluated from Exercise 2.2:

$$\sum (\cos \omega)^2 = \frac{n}{2} \{1 + D_n(2\omega) \cos(n-1)\omega\}, \quad \sum \cos \omega \sin \omega = \frac{n}{2} D_n(2\omega) \sin(n-1)\omega,$$

$$\sum (\sin \omega)^2 = \frac{n}{2} \{1 - D_n(2\omega) \cos(n-1)\omega\}, \quad \sum \cos \omega = n D_n(\omega) \cos((n-1)\omega/2),$$

$$\sum \sin \omega = n D_n(\omega) \sin((n-1)\omega/2).$$

If $\omega$ is not too close to zero and $n$ is large, $D_n(2\omega)$ is negligible. (And if $\omega$ is a nonzero Fourier frequency, then $D_n(2\omega)$ is exactly zero.) Thus, we obtain the approximations

$$\sum \cos^2 \omega t = \frac{n}{2}, \quad \sum \sin^2 \omega t = \frac{n}{2}, \quad \sum \cos \omega \sin \omega = 0, \quad \sum \cos \omega = 0, \quad \sum \sin \omega = 0.$$

The normal equations then yield the approximate least squares estimators

$$\mu = \frac{1}{n} \sum x_t, \quad A = \frac{2}{n} \sum x_t \cos \omega t, \quad B = \frac{2}{n} \sum x_t \sin \omega t.$$
Note that \( A \) and \( B \) are closely related to the DFT of the data, \( J(\omega) \). Specifically, \( A = 2 \Re [J(\omega)] \) and \( B = -2 \Im [J(\omega)] \). If \( \omega = \omega_j \) is a Fourier frequency, then \( A \) and \( B \) are the cosine and sine sums \( A_j, B_j \) introduced in Part 1.

**Decomposing the Sum of Squares**

From the harmonic representation

\[
x_i = \sum_{j=0}^{n-1} J_j \exp(i \omega_j t),
\]

\( x_i \) is a sum of (complex) harmonic components at frequencies \( \omega_0, \ldots, \omega_{n-1} \). The component at \( \omega_j \) is the series (or vector) \( J_j \exp(i \omega_j t), t = 0, \ldots, n-1 \). This vector has squared modulus (sum of squares)

\[
\sum_{t=0}^{n-1} |J_j \exp(i \omega_j t)|^2 = \sum_{t=0}^{n-1} |J_j|^2 = n |J_j|^2.
\]

The sum of squares of the data vector \( \{x_t\}_{t=0}^{n-1} \) has a nice representation in terms of the sums of squares of the harmonic components of \( \{x_t\} \):

\[
\sum_{t=0}^{n-1} |x_t|^2 = \sum_{t=0}^{n-1} |\sum_{j=0}^{n-1} J_j \exp(i \omega_j t)|^2 = \sum_{t=0}^{n-1} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} J_j J_k^* \exp(i \omega_{j-k} t)
\]

\[
= \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} J_j J_k^* \sum_{t=0}^{n-1} \exp(i \omega_{j-k} t) = \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} J_j J_k^* n \cdot \{j = k\} = n \sum_{j=0}^{n-1} |J_j|^2.
\]

Thus, the sum of squares of \( \{x_t\} \) is decomposed as the sum of the sums of squares \( n |J_j|^2 \) of the harmonic components of \( \{x_t\} \).

**The Periodogram**

The periodogram was invented by Schuster in 1898 for detecting periodic components in time series data. It is defined by

\[
I_j = \frac{n}{2\pi} |J_j|^2, \quad j = 0, \ldots, n-1.
\]

More generally, the periodogram is defined for all frequencies \( \omega \) by \( I(\omega) = \frac{n}{2\pi} |J(\omega)|^2 \).
Note that \( I_j = I(\omega_j) \).

Note also that the sum of squares of \( \{x_t\} \) is

\[
\sum_{t=0}^{n-1} |x_t|^2 = 2\pi \sum_{j=0}^{n-1} I_j,
\]

a constant multiple of the sum of the periodogram values.

The periodogram ordinate \( I_j \) is the contribution to the total sum of squares of \( \{x_t\} \) due to the harmonic component at frequency \( \omega_j \). If \( x_t = A \cos(\omega_j t) + B \sin(\omega_j t) \), a pure sinusoid oscillating at an exact Fourier frequency, then all periodogram ordinates will be zero except for \( I_j \). In general, a peak in the periodogram at a given frequency indicates a strong harmonic component in the data \( \{x_t\} \) at that frequency. In the model (1), if the true frequency \( \omega \) is unknown, it can be estimated by maximizing \( I(\lambda) \) over all frequencies. To reduce the computational expense of the procedure, we may wish to restrict the search to Fourier frequencies \( \omega_j \).

If used wisely, the periodogram can be a very useful tool. If used naively, however, it can lead the data analyst to find periodic components which do not really exist in the underlying data-generating mechanism. For example, in model (1), even if \( A \) and \( B \) are zero, the periodogram will typically exhibit several strong peaks. Thus, until we can derive the distribution of the periodogram under the null hypothesis where there are no underlying periodic components, we should be careful not to declare that we have found a "true" cycle. Many of the past attempts at detecting "business cycles" in economic data are flawed by their failure to account for random variation in the periodogram.

**Smooth Functions**

Qualitatively speaking, a sequence \( x_0, \ldots, x_{n-1} \) is smooth if its DFT is large only at low frequencies. This is true since \( \exp(i\omega_j t) \) for \( j \) small is a slowly varying (hence smooth) function, and since any linear combination of smooth functions is smooth. On the other hand, if \( j \) is large, \( \exp(i\omega_j t) \) oscillates rapidly and hence is not smooth. No smooth sequence can contain such components with large amplitude. A quantitative measure of roughness is the (circular) roughness coefficient,

\[
\sum_{t=0}^{n-1} (x_t - x_{t-1})^2 / \sum_{t=0}^{n-1} x_t^2.
\]
To see how the coefficient measures roughness, we use frequency domain methods. Since
\[ x_t = \sum_{j=0}^{n-1} J_j \exp(i \omega_j t), \]
we have
\[ x_t - x_{t-1} = \sum_{j=0}^{n-1} J_j (\exp(i \omega_j t) - \exp(i \omega_j (t-1))) = \sum_{j=0}^{n-1} J_j \exp(i \omega_j t) \{1 - \exp(-i \omega_j)\}, \]
resulting in
\[ \frac{\sum_{j=0}^{n-1} |1 - \exp(-i \omega_j)|^2 |J_j|^2}{\sum_{j=0}^{n-1} |J_j|^2} = \frac{4 \sum_{j=0}^{n-1} (\sin \omega_j/2)^2 |J_j|^2}{\sum_{j=0}^{n-1} |J_j|^2}. \]
(The proof of the last equality is left as an exercise.) Thus, for a fixed value of the denominator, the roughness coefficient is reduced by increasing the low frequency terms at the expense of the high frequency terms.

**Aliasing**

Suppose that our time series is an exact cosine wave, \( x_t = \cos \omega t \). The frequency \( \omega \) may be any real number, but the fact that we only have observations at the equally spaced time points \( t = 0, \ldots, n - 1 \) implies that certain frequencies will be indistinguishable from each other (aliases). In fact, the largest frequency (i.e., the most rapid oscillation) we can observe is \( \omega = \pi \), which corresponds to the series \( x_t = \cos \pi t = (-1)^t \). For any larger value of \( \omega \), the series \( x_t = \cos \omega t \) is indistinguishable from the series \( x_t = \cos \omega' t \ (t = 0, \ldots, n - 1) \) for some \( \omega' \) such that \( 0 \leq \omega' \leq \pi \). (Example given below.) The frequencies \( \omega \) and \( \omega' \) are said to be aliases of each other, and \( \omega' \) is the principal alias. For example, suppose that \( \omega \) is between \( \pi \) and \( 2\pi \). Then we can define \( \omega' = 2\pi - \omega \). Note that \( \omega' \) is between 0 and \( \pi \), and
\[ \cos \omega' t = \cos(2\pi t - \omega' t) = \cos(2\pi t) \cos(-\omega t) - \sin 2\pi t \sin(-\omega t) = \cos \omega t. \]
Thus, the frequencies \( \omega' \) and \( \omega \) are indistinguishable, as long as the time spacing between observations (the "sampling rate") is one time unit. Even if the series really is \( \cos \omega' t \), it looks just like the series \( \cos \omega t \) when observed at equally spaced intervals of one time unit, \( t = 0, 1, \ldots, n - 1 \). To see the
difference, we would have to increase the sampling rate.

As stated above, the maximum observable frequency is $\pi$ radians per unit time. It is called the **folding frequency** (or Nyquist frequency). All higher frequencies are "folded" down into the interval $[0, \pi]$. A sinusoid at the Nyquist frequency executes 1 complete cycle for every two units of time (that is, one half cycle per unit time, or $2\pi(1/2) = \pi$ radians per unit time). Thus, for a frequency to be observable (i.e., in the range of principal aliases), at least two samples (observations) per cycle are required.