8 Chapter 8: Estimating Spectral Densities

8.1 Regression on sinusoidal components

The simplest form of spectral analysis consists of regression on a periodic component:

$$Y_t = A\cos\omega t + B\sin\omega t + C + \varepsilon_t, \qquad 1 \le t \le T, \tag{8.1}$$

where $\varepsilon_t \sim WN(0, \sigma^2)$. Without loss of generality, we assume $0 \le \omega \le \pi$. In fact, for discrete data frequencies outside this range are **aliased** into this range. Indeed, let $\Phi = N\pi + \omega$ where N is integer and $0 \in [0, \pi]$. If N is even then

$$\cos(\Phi t) = cos(\omega t).$$

If N is odd then

$$\cos(\Phi t) = cos([(2k-1)\pi + \omega]t) = cos([\omega - \pi]t) = cos([\pi - \omega]t),$$

Φ is aliased to some frequency in the interval $[0, \pi]$.

Hence, a sampled sinusoid with frequency smaller than 0, appears to always coincide with a sinusoid with frequency from $[0, \pi]$.

Remark. The term $A\cos\omega t + B\sin\omega t$ is a periodic function with period $\frac{2\pi}{\omega}$. The period $2\pi/\omega$ represents the number of time units that it takes for the function to take the same value again, i.e. to complete a cycle. The frequency, measured in cycles per time unit, is given by the inverse $\frac{\omega}{2\pi}$. The **angular frequency**, measured in radians per time unit, is given by ω. Because of its convenience, the angular frequency ω will be used to describe periodicity of the function, and its name is shortened to frequency, when there is no danger of confusion.

Example. Consider the monthly data that exhibit a 12-month seasonality. Hence, the period $\frac{2\pi}{\omega}$ is equal to 12, which implies that the angular frequency $\omega = \frac{\pi}{6}$. Also the frequency, measured in cycles per time unit, is given by the inverse $\frac{\omega}{2\pi} = 1/12 \approx 0.08$.

We may rewrite this model in a standard matrix form

$$Y = X\beta + \varepsilon,$$

where $Y = (Y_1, \ldots, Y_T)'$ is a vector of observations, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$ is a vector of errors, $\beta = (A, B, C)'$ is a vector of model parameters; X is the $T \times 3-$ matrix

$$X = \begin{pmatrix} \cos\omega & \sin\omega & 1 \\ \cos 2\omega & \sin 2\omega & 1 \\ \vdots & \vdots & \ddots \\ \cos T\omega & \sin T\omega & 1 \end{pmatrix},$$

Then the estimates of parameters $\hat{\beta}$ are given by the ordinary least squares (OLS) method as

$$\hat{\beta} = \begin{pmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{pmatrix} = \left(X^T X\right)^{-1} X^T Y = \left(X^T X\right)^{-1} \begin{pmatrix} \sum_t Y_t \cos\omega t \\ \sum_t Y_t \sin\omega t \\ \sum_t Y_t \end{pmatrix}.$$

The estimate $\hat{\beta}$ is unbiased with a variance-covariance matrix given by the usual formula $(X^T X)^{-1}\sigma^2$.

These formulas take a much simpler form if ω is one of the Fourier frequencies, defined by

$$\omega_j = 2\pi j/T, \quad 0 \le j \le \frac{T}{2}.$$

We may note the following elementary trigonometric identities:

$$\sum_{t=1}^{T} \cos\omega_j t = \sum_{t=1}^{T} \sin\omega_j t = 0 \qquad (j \neq 0), \tag{8.2}$$

$$\sum_{t=1}^{T} \cos\omega_j t \cos\omega_k t = \begin{cases} \frac{T}{2} & \text{if } \; j = k \neq 0 \; \text{ or } \; j = k \neq \frac{T}{2} \\ T & \text{if } \; j = k = 0 \; \text{ or } \; j = k = \frac{T}{2}, \\ 0 & \text{if } \; j \neq k, \end{cases} \tag{8.3}$$

$$\sum_{t=1}^{T} \sin\omega_j t \sin\omega_k t = \begin{cases} \frac{T}{2} & \text{if } j = k \neq 0 \; \text{ or } \; j = k \neq \frac{T}{2}, \\ 0 & \text{otherwise,} \end{cases} \tag{8.4}$$

$$\sum_{t=1}^{T} \cos\omega_j t \sin\omega_k t = 0 \qquad (\forall j, k \epsilon Z). \tag{8.5}$$

(Exercise: verify these equations. Hint: $\sum_{t=0}^{T-1} e^{i\omega_j t} = \sum_{t=0}^{T-1} e^{\frac{i2\pi j t}{T}} = \frac{1-e^{i2\pi j}}{1-e^{i2\pi j/T}}$ and $e^{i2\pi j} = \cos(2\pi j) + i\sin(2\pi j) = 1$.)

For any Fourier frequency ω_j (except of the cases $j = 0,\ j = T/2$ that produce similar but slightly different formulas), we obtain

$$X^T X = \begin{pmatrix} \frac{T}{2} & 0 & 0 \\ 0 & \frac{T}{2} & 0 \\ 0 & 0 & T \end{pmatrix},$$

Hence,

$$\begin{aligned} \hat{A} &= \frac{2}{T} \sum_t Y_t \cos \omega_j t, \\ \hat{B} &= \frac{2}{T} \sum_t Y_t \sin \omega_j t, \\ \hat{C} &= \overline{Y} = \frac{1}{T} \sum_t Y_t, \end{aligned}$$

are uncorrelated estimates of A, B and C with variances $2\sigma^2/T, 2\sigma^2/T, \sigma^2/T$, respectively.

A suitable way of testing the significance of the sinusoidal component with frequency ω_j is using its contribution to the sum of squares

$$R_T(\omega_j) = \frac{T}{2} \left(\hat{A}^2 + \hat{B}^2 \right). \tag{8.6}$$

If the $\varepsilon_t \sim N(0, \sigma^2)$, then it follows that $\hat{A}$ and $\hat{B}$ are also independent normal each with variance $2\sigma^2/T$, so under the null hypothesis $A = B = 0$ we find that

$$\frac{R_T(\omega_j)}{\sigma^2} \sim \chi_2^2 \tag{8.7}$$

or equivalently that $R_T(\omega_j)/(2\sigma^2)$ has an exponential distribution with mean 1.

The above theory is easily extended to simultaneous estimation of several periodic components. In particular, if we consider estimation of sinusoidal terms at k Fourier frequencies $\omega_{j_1}, \ldots, \omega_{j_k}$, the corresponding columns of the X matrix are orthogonal. This implies that the point estimates of the coefficients are the same when all k components are estimated simultaneously and when they are estimated one at a time. Also that the parameter estimates and hence the R_T statistics are uncorrelated.

Under the null hypothesis that all the A and B coefficients are 0, we have the following result:

(*) The k test normalized statistics $R_T(\omega_{j_1})/(\sigma^2), \ldots, R_T(\omega_{j_k})/(\sigma^2)$ are independent χ^2-distributed random variables with 2 degrees of freedom.

This is an exact result for all T if $\varepsilon_t \sim N(0, \sigma^2)$. It is also valid as an approximation for large T if the $\{\varepsilon_t\}$ is non-normal white noise. This is because the Central Limit Theorem guarantees that the parameter estimates are approximately independent normal for large T.

8.2 The Periodogram

The foregoing discussion turns out to be very useful in studying some of the fundamental properties of the periodogram. Recall that the periodogram was defined as

$$I_T(\omega) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} Y_t e^{i\omega t}\right|^2, \tag{8.8}$$

where it was stated (without any proof) that it is an approximately unbiased estimator of the spectral density f. In this section we shall provide an informal proof of this, and derive some related statistical properties along the way.

In terms of the $R_T(\omega)$ statistic defined above, the periodogram is

$$I_T(\omega_j) = \frac{R_T(\omega_j)}{4\pi}. \tag{8.9}$$

(Indeed, take into account the relationship between (8.8), $\hat{A}^2$ and $\hat{B}^2$ as well as orthogonality of Fourier basis.) By the property (*), if the process is Gaussian white noise then the $I_T(\omega_j)$ at Fourier frequencies $\{\omega_j,\ 1 \le j < T/2\}$ is a multiple of χ^2-random variable with two degrees of freedom, i.e.

$$I_T(\omega_j) \sim \frac{\sigma^2 \chi_2^2}{4\pi} = \frac{f_{WN}(\omega_j)}{2}\chi_2^2.$$

(Recall that the spectral density of white noise $f_{WN}(\omega)$ is $\sigma^2/2\pi$.)

This result is exact if $\varepsilon_t \sim N(0, \sigma^2)$, and approximate for large T if ε is non-normal white noise.

We can state a more general result as follows.

Theorem. Suppose $Y_t = \sum_{r=0}^{\infty} c_r \varepsilon_{t-r}$ is a linear process where $\varepsilon_t \sim WN(0, \sigma^2)$. Suppose this process Y_t is weakly stationary with a spectral density

$$f(\omega) = \frac{\sigma^2}{2\pi}\left|C\left(e^{i\omega}\right)\right|^2, \qquad C(z) = \sum_{r=0}^{\infty} c_r z^r. \tag{8.10}$$

Then the periodogram ordinates $\{I_T(\omega_j),\ 1 \le j < T/2\}$, are approximately independent and exponentially distributed, with means $\{f(\omega_j),\ 1 \le j < T/2\}$, i.e.

$$I_T(\omega_j) \sim \frac{f(\omega_j)}{2}\chi_2^2.$$

Heuristic proof. Let us write

$$\begin{aligned} \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T} Y_t e^{i\omega t} &= \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\sum_{r=0}^{\infty} c_r \varepsilon_{t-r} e^{i\omega t} \\ &= \frac{1}{\sqrt{2\pi T}}\sum_{r=0}^{\infty} c_r e^{i\omega r}\sum_{t=1}^{T} \varepsilon_{t-r} e^{i\omega(t-r)} \\ &= \frac{1}{\sqrt{2\pi T}}\sum_{r=0}^{\infty} c_r e^{i\omega r}\sum_{u=1-r}^{T-r} \varepsilon_u e^{i\omega u}. \end{aligned} \tag{8.11}$$

For large T and fixed r, we may approximate

$$\sum_{u=1-r}^{T-r} \varepsilon_u e^{i\omega u} \approx \sum_{u=1}^{T} \varepsilon_u e^{i\omega u} \tag{8.12}$$

essentially because we are adding or removing a total of r terms and these are negligible compared with the total length of the sum T.

If we accept the approximation 8.12, then we get

$$\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T} Y_t e^{i\omega t} \approx C(e^{i\omega})\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T} \varepsilon_t e^{i\omega t} \tag{8.13}$$

and hence, on taking squared absolute values of both sides,

$$I_{T,Y}(\omega) \approx \left|C(e^{i\omega})\right|^2 I_{T,\varepsilon}(\omega) \tag{8.14}$$

with obvious notation: $I_{T,Y}$ and $I_{T,\varepsilon}$ are the periodograms of the Y and ε processes, respectively.

In combination with the result (*), 8.14 shows that the periodogram ordinates of Y, evaluated at Fourier frequencies, are approximately independent. Moreover,

$$E\left\{I_{T,Y}(\omega_j)\right\} \approx \left|C(e^{i\omega_j})\right|^2 E\left\{I_{T,\varepsilon}(\omega_j)\right\} = f(\omega_j). \tag{8.15}$$

□

This theorem is the central result of spectral estimation theory.

However, this result also points to some undesirable features of the periodogram:

- $I_T(\omega)$ for a fixed ω is not a consistent estimate of $f(\omega)$, since

  $$I_T(\omega_j) \sim \frac{f(\omega_j)}{2}\chi_2^2.$$

  Therefore variance does not tend to 0 as $T \to \infty$.

- Also, the independence of periodogram ordinates at different Fourier frequencies suggests that the sample periodogram, plotted as a function of ω, will be extremely irregular.

Smoothing The idea behind smoothing is to take weighted averages over neighboring frequencies in order to reduce the variability associated with individual periodogram values. However, such an operation necessarily introduces some bias into the estimation procedure. Theoretical studies focus on the amount of smoothing that is required to obtain an optimum trade-off between bias and variance. In practice, this usually means that choice of a kernel and amount of smoothing is somewhat subjective.

The main form of smoothed estimator is given by

$$\hat{f}(\lambda) = \int_{-\pi}^{\pi} \frac{1}{h} K\left(\frac{\omega - \lambda}{h}\right) I_T(\omega) d\omega \tag{8.16}$$

where $IT(\cdot)$is the periodogram based on T observations, $K()$ is a kernel function and h is the *bandwidth*. We usually take $K(\cdot)$ to be a non-negative function, symmetric about 0, and integrating to 1. Thus, any symmetric density, such as the normal, will work. In practice, however, it is more usual to take a kernel of finite range, such as the *Epanechnikov* kernel

$$K(x) = \frac{3}{4\sqrt{5}}\left(1 - \frac{t^2}{5}\right), \qquad -\sqrt{5} \leq t \leq \sqrt{5}, \tag{8.17}$$

that is 0 outside $[-\sqrt{5}, \sqrt{5}]$. This choice of kernel function has some optimality properties. However, in practice this optimality is less important than the choice of bandwidth h, which effectively controls the range over which the periodogram is smoothed.

There are some additional difficulties with the performance of the sample periodogram in the presence of a sinusoidal variation whose frequency is not one of the Fourier frequencies. This effect is known as *leakage*. The reason of "leakage" is that we always consider a truncated periodogram 8.8. Truncation implicitly assumes that the time series is periodic with period T, which, of course, is not always true. So we artificially introduce non-existent periodicities into the estimated spectrum, i.e. cause "leakage" of the spectrum. (If time series is perfectly periodic over T then there is no "leakage".) The "leakage" can be treated using an operation of *tapering* on the periodogram, i.e. by choosing appropriate periodogram windows.

Remark. When we work with periodograms, we loose all phase (relative location/time origin) information: the periodogram will be the same if all the data were circularly rotated to a new time origin, i.e. the observed data are treated as perfectly periodic.

8.3 Examples

The monthly productions of chocolate confectionery in Australia (in tonnes) July 1957 - Aug 1995 are very seasonal as we noticed in the previous lectures.

Here "spans" control the degree of smoothing. Alternatively you can choose different type of kernels.

The x-axis corresponds to $\frac{\omega}{2\pi}$. Notice that we have a very large peak at an approximate frequency 0.08. Thus, $\frac{\omega}{2\pi} \approx 0.08$, which implies $\frac{2\pi j}{12}\frac{1}{2\pi} = \frac{j}{12} \approx 0.08$ and $j \approx 0.96$, i.e. we have one cycle per year. That is exactly what we got in our previous lectures.

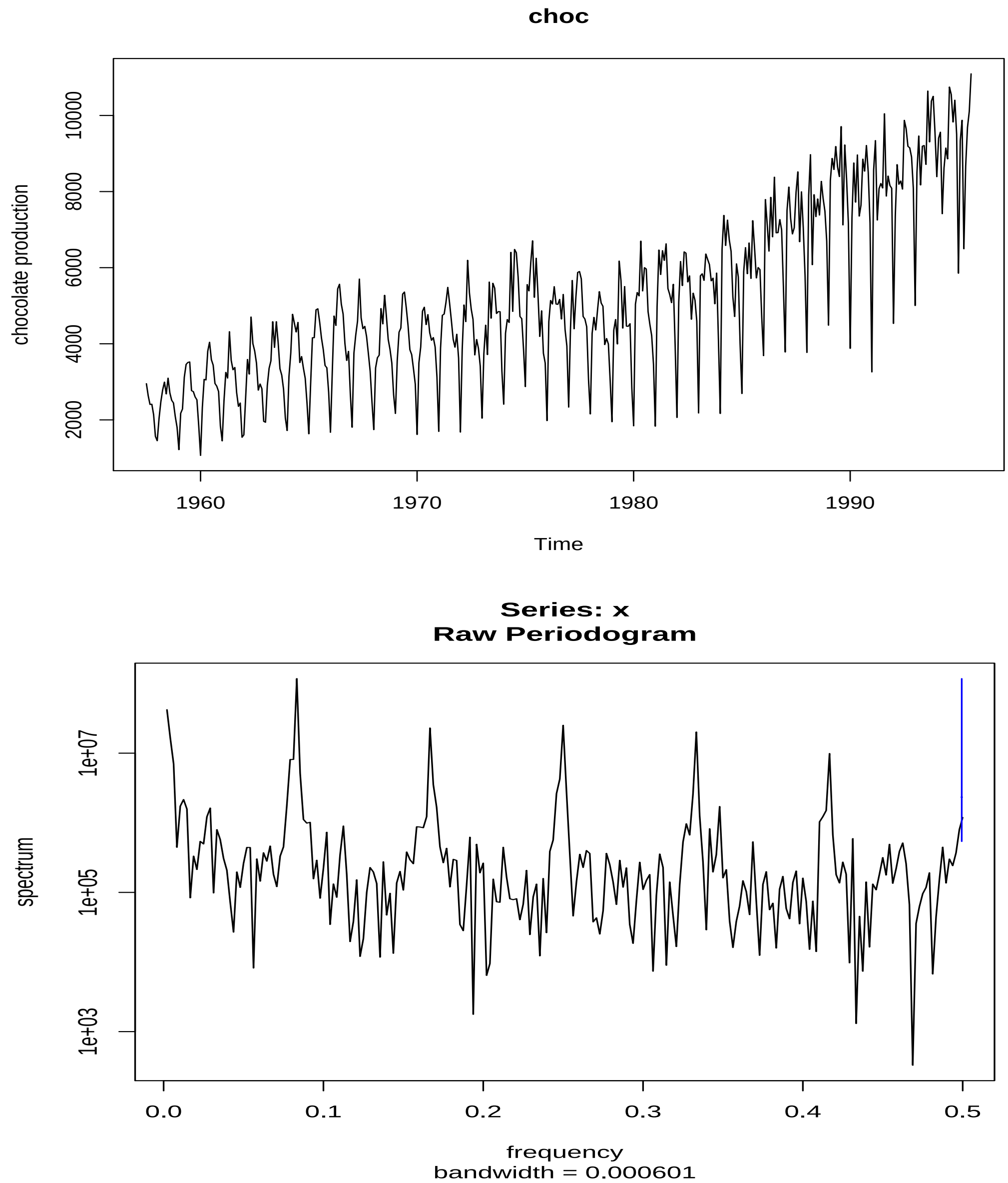

Figure 8.1: Time Series Plot and Raw Periodogram of the chocolate production in Australia, 1957 - 1995.

```
> par(mfrow=c(2,2))
> spectrum(choc, spans=c(3,3))
```

```
> spectrum(choc, spans=c(5,5))
> spectrum(choc, spans=c(7,7))
> spectrum(choc, spans=c(21,21))
```

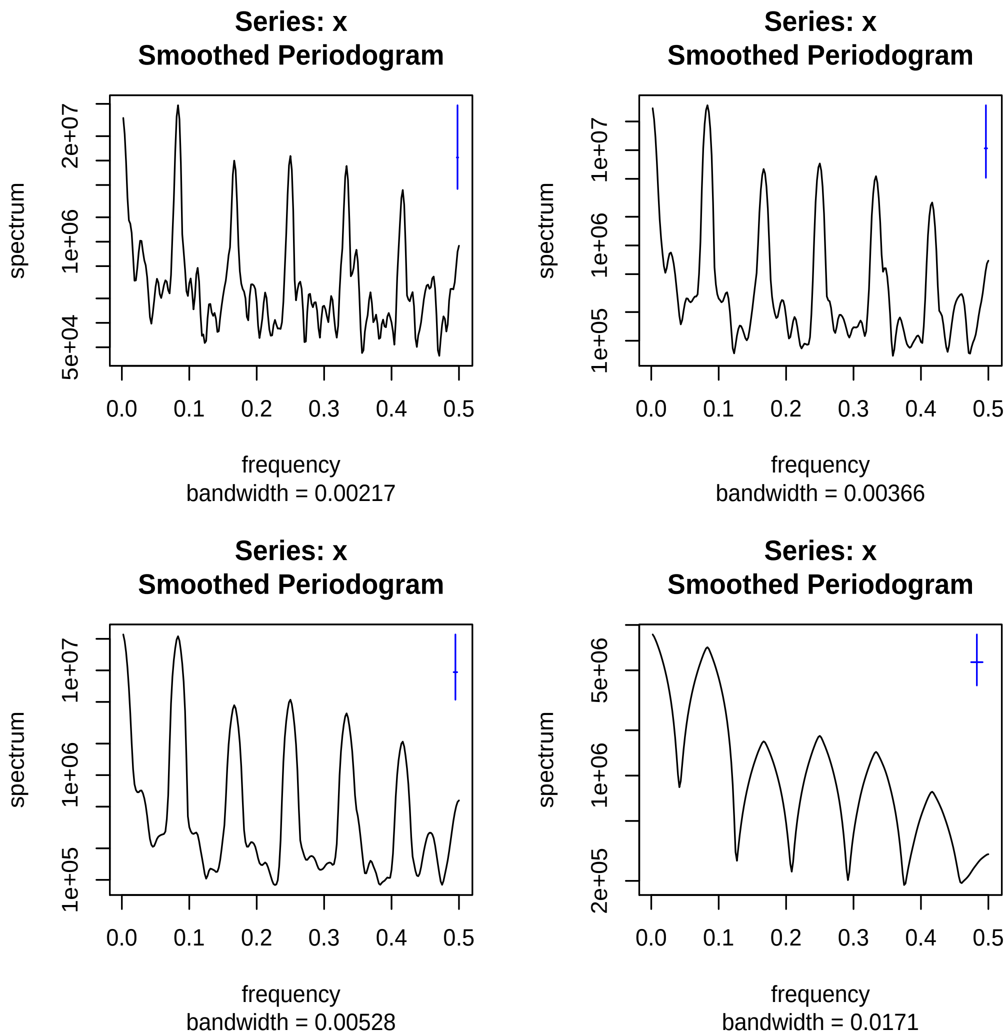

Figure 8.2: Smoothed Periodogram of Chocolate production in Australia, 1957 - 1995.

Notice that you can use the confidence band in the upper right corner to get an approximate idea how significant the peak is.

Another method of testing the reality of a peak is to look at its harmonics. It is extremely unlikely that a true cycle will be shaped perfectly as a sine curve and at least the first few harmonics will show up as well. For example, if we have monthly data with the annual seasonality (12 months period) then it will almost certainly the periodogram will not look as a perfect sine function. In contrast, the peaks at 6-, 4-, 3- months and possibly others will show up and will be also of importance if 12-month peak is important (see Granger, 1964 for more discussion). It is interesting that this is exactly the pattern that we see in Fig. 8.2-8.3.

Also we can try to approximate our data with an AR model and then plot the approximating periodogram of an AR model.

```
> spectrum(choc, method="ar")
```

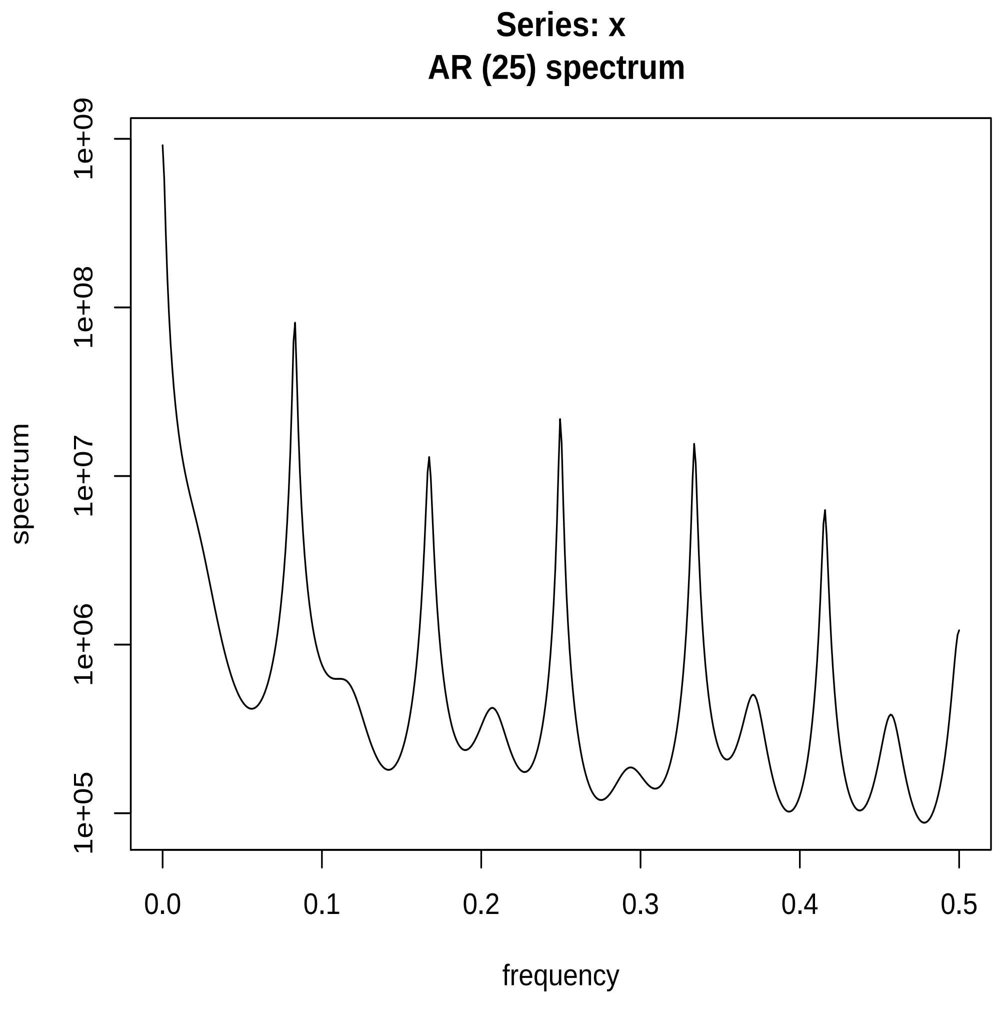

Figure 8.3: AR Approximated Periodogram of Chocolate production in Australia, 1957 - 1995.

8.4 Example: Yearly Sunspot Data 1700-1988

One of the classical examples is the sunspot observations. The earliest surviving record of sunspot dates from the 364 B.C., according a star catalogue by Chinese astronomer Gan De. Let us we take a sample of annual sunspot observations from 1700 to 1988 (see Figure 8.4). It is well-known that the sunspot populations rise and fall on an irregular cycle of 11 years, and this is exactly what we see in Figure 8.4.

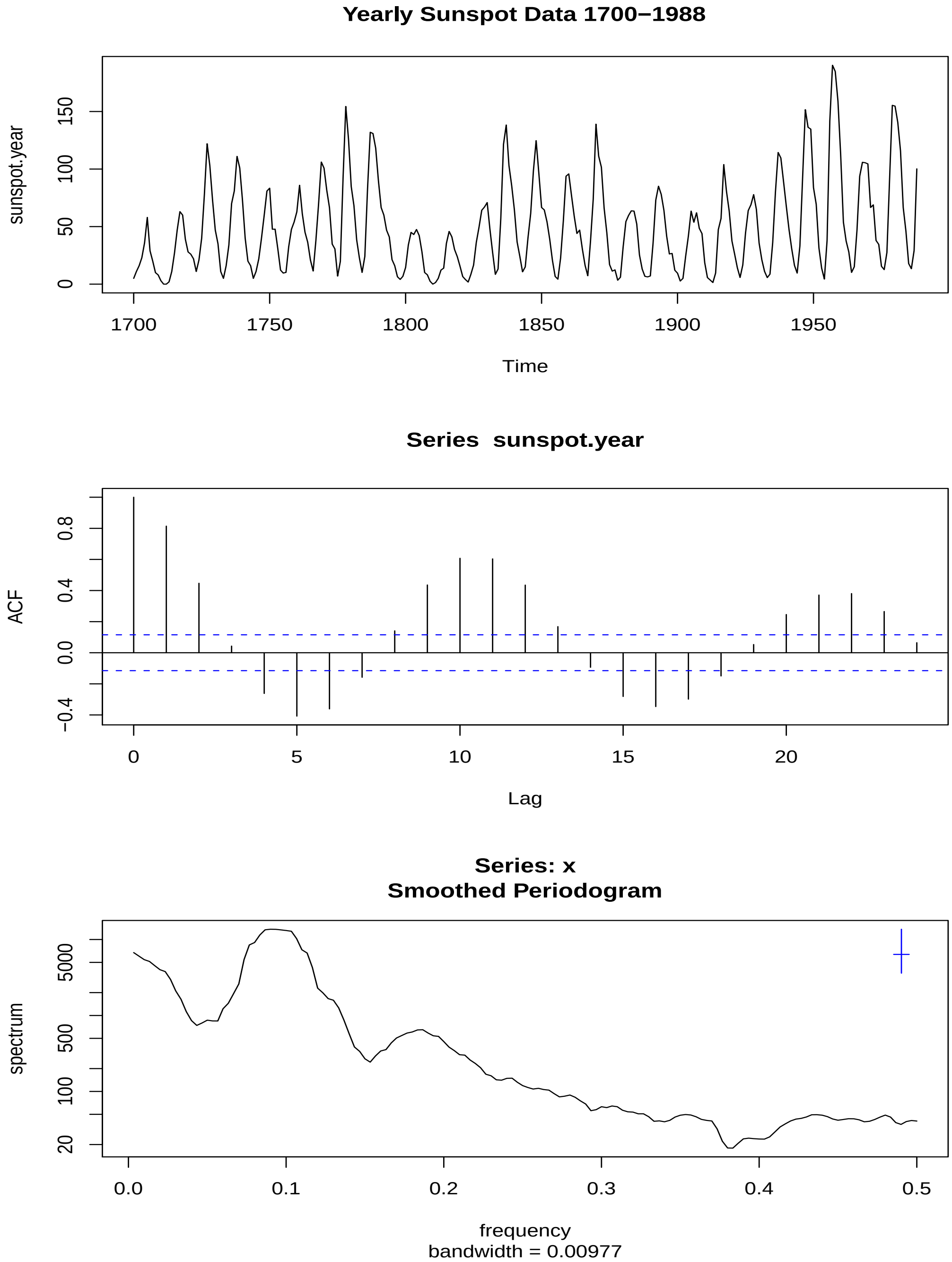

Figure 8.4: Sunspot Data 1700-1988