Lecture 3: Spectral Analysis*

Any covariance stationary process has both a *time domain* representation and a *spectrum domain* representation. So far, our analysis is in the time domain as we represent a time series $\{x_t\}$ in terms of past values of innovations and investigate the dependence of x at distinct time. In some cases, a spectrum-domain representation is more convenient in describing a process. To transform a time-domain representation to a spectrum-domain representation, we use the Fourier transform.

1 Fourier Transforms

Let ω denote the *frequency* $(-\pi < \omega < \pi)$, and let T denote the *period*: the minimum time that it takes the wave to go through a whole cycle, and we have $T = 2\pi/\omega$. Given any integer number z, we have $x(t) = x(t + zT)$. Finally, we will let ϕ denote the *phase*: the amount that a wave is shifted.

Given a time series $\{x_t\}$, its Fourier transformation is:

$$x(\omega) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} e^{-it\omega} x(t) \tag{1}$$

and the inverse Fourier transform is:

$$x(t) = \int_{-\pi}^{\pi} e^{it\omega} x(\omega) d\omega \tag{2}$$

2 Spectrum

Recall that the autocovariance function for a zero-mean stationary process $\{x_t\}$ is defined as:

$$\gamma_x(h) = E(x_t x_{t-h})$$

and it serves to characterize the time series $\{x_t\}$. The *spectrum* of $\{x\}$ is defined to be the Fourier transform of $\gamma_x(h)$,

$$S_x(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\omega} \gamma_x(h) \tag{3}$$

Recall that the autocovariance generating function is $g_x(z) = \sum_{h=-\infty}^{\infty} \gamma_x(h) z^h$, if we let $z = e^{-i\omega}$, then the spectrum is just the autocovariance generating function divided by 2π. In (3), if we take $\omega = 0$, we see that

$$\sum_{h=-\infty}^{\infty} \gamma_x(h) = 2\pi S_x(0),$$

which tells that the sum of autocorrelations equals the spectrum at zero multiplied by 2π. Using the identity

$$e^{i\phi} = \cos\phi + i\sin\phi,$$

we can also write (3) as

$$S_x(\omega) = \frac{1}{2\pi}\left[\gamma_0 + 2\sum_{h=1}^{\infty}\gamma_x(h)\cos(h\omega)\right]. \tag{4}$$

Note that since $\cos(\omega) = \cos(-\omega)$, and $\gamma_x(h) = \gamma_x(-h)$, the spectrum is symmetric about zero. Also the cosine function is periodic with period 2π, therefore, for spectral analysis, we only need to find the spectrum for $\omega \in [0, \pi]$. Now if we know $\gamma_x(h)$, we can compute its spectrum using (4), and if we know the spectrum $S_x(\omega)$, we can compute $\gamma_x(h)$ using the inverse Fourier transform:

$$\gamma_x(h) = \int_{-\pi}^{\pi} e^{i\omega h} S_x(\omega) d\omega \tag{5}$$

Let $h = 0$, then (5) gives the variance of $\{x_t\}$

$$\gamma_x(0) = \int_{-\pi}^{\pi} S_x(\omega) d\omega.$$

So the variance of $\{x_t\}$ is just the sum of the spectrum over all frequencies $-\pi < \omega < \pi$. Therefore we can see that the spectrum function $S_x(\omega)$ decomposes the variance into components contributed from each frequency. In other words, we can use spectrum to find the importance of cycles of different frequencies.

If we normalize the spectrum $S_x(\omega)$ by dividing $\gamma_x(0)$, we get the Fourier transform of the autocorrelation function $\rho_x(h)$,

$$f(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} e^{-ih\omega}\rho_x(h) \tag{6}$$

The autocorrelation functions can be generated from $f(\omega)$ using the inverse transform

$$\rho_x(h) = \int_{-\pi}^{\pi} e^{i\omega h} f_x(\omega) d\omega \tag{7}$$

Again, let $h = 0$, (7) gives

$$1 = \int_{-\pi}^{\pi} f_x(\omega) d\omega$$

Note that $f(\omega)$ is positive and integrate to one, just like a probability distribution density, so we call it *spectral density.*

Example 1 (spectral density of white noise) Let $\epsilon \sim \text{WN}(0, \sigma_\epsilon^2)$. We have $\gamma_\epsilon(0) = \sigma_\epsilon^2$ and $\gamma_\epsilon(h) = 0$ for $h \neq 0$. Using (3) and (6), we can compute

$$S_\epsilon(\omega) = \frac{1}{2\pi}\gamma_\epsilon(0) = \frac{1}{2\pi}\sigma_\epsilon^2.$$

Divide it by $\gamma_\epsilon(0)$, we have

$$f_x(\epsilon) = \frac{1}{2\pi}.$$

So the spectral density is uniform over $[-\pi, \pi]$, i.e., every frequency has equal contribution to the variance.

3 Spectrum of Filtered Process

Considering that the spectrum of a white noise process is so simple, we may want to know if we could make use it for a more complicated process, say,

$$x_t = \sum_{k=-\infty}^{\infty} \theta_k \epsilon_{t-k} = \theta(L)\epsilon_t.$$

We call this process a two-sided moving average process. Then what is the relationship between $S_x(\omega)$ and $S_\epsilon(\omega)$? The general solution is given in the following statement.

Proposition 1 *If $\{x_t\}$ is a zero mean stationary process with spectrum function $S_x(\omega)$, and $\{y_t\}$ is the process*

$$y_t = \sum_{k=-\infty}^{\infty} \theta_k x_{t-k} = \theta(L)x_t$$

where θ is absolutely summable, then

$$S_y(\omega) = \left| \sum_{k=-\infty}^{\infty} \theta_k e^{-ik\omega} \right|^2 S_x(\omega) = \left|\theta(e^{-i\omega})\right|^2 S_x(\omega).$$

Proof: We start from the autocovarinace function of y,

$$\begin{aligned}
\gamma_y(h) &= E(y_t y_{t-h}) \\
&= E\left(\sum_{j=-\infty}^{\infty} \theta_j x_{t-j} \sum_{k=-\infty}^{\infty} \theta_k x_{t-h-k} \right) \\
&= \sum_{j,k=-\infty}^{\infty} \theta_j \theta_k E(x_{t-j} x_{t-h-k}) \\
&= \sum_{j,k=-\infty}^{\infty} \theta_j \theta_k \gamma_x(h+k-j)
\end{aligned}$$

Next, consider the spectrum of y,

$$\begin{aligned}
S_y(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\omega} \gamma_y(h) \\
&= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\omega} \sum_{j,k=-\infty}^{\infty} \theta_j \theta_k \gamma_x(h+k-j)
\end{aligned}$$

(Let $l = h+k-j$ and note that $S_x(\omega) = \frac{1}{2\pi}\sum_{l=-\infty}^{\infty} e^{-il\omega}\gamma_x(l)$, so we want to construct such a term and see what are the remainings.)

$$\begin{aligned}
S_y(\omega) &= \sum_{j=-\infty}^{\infty} e^{-ij\omega}\theta_j \sum_{k=-\infty}^{\infty} e^{ik\omega}\theta_k \left(\frac{1}{2\pi} \sum_{l=-\infty}^{\infty} e^{-il\omega}\gamma_x(l) \right) \\
&= \theta(e^{-i\omega})\theta(e^{i\omega}) S_x(\omega) \\
&= \theta(e^{-i\omega})\overline{\theta(e^{-i\omega})} S_x(\omega) \\
&= \left|\theta(e^{-i\omega})\right|^2 S_x(\omega)
\end{aligned}$$

Example 2 To apply this results, first consider the problem of computing an MA(1) process,

$$x_t = \epsilon_t + \theta\epsilon_{t-1} = (1 + \theta L)\epsilon_t.$$

In this problem,

$$\theta(e^{-i\omega}) = 1 + \theta e^{-i\omega},$$

thus

$$\begin{aligned} \left|\theta(e^{-i\omega})\right|^2 &= (1 + \theta e^{-i\omega})(1 + \theta e^{i\omega}) \\ &= 1 + \theta^2 + \theta(e^{-i\omega} + e^{i\omega}) \end{aligned}$$

Therefore,

$$\begin{aligned} S_x(\omega) &= \left|\theta(e^{-i\omega})\right|^2 S_\epsilon(\omega) \\ &= \frac{1}{2\pi}[1 + \theta^2 + \theta(e^{-i\omega} + e^{i\omega})]\sigma_\epsilon^2 \end{aligned}$$

We can verify this result by using the spectrum to compute the autocovarinace function, say, $\gamma_x(1)$. Using (5).

$$\begin{aligned} \gamma_x(1) &= \int_{-\pi}^{\pi} e^{i\omega} S_x(\omega) d\omega \\ &= \frac{1}{2\pi}\sigma_\epsilon^2 \int_{-\pi}^{\pi} e^{i\omega}[1 + \theta^2 + \theta(e^{-i\omega} + e^{i\omega})]d\omega \\ &= \frac{1}{2\pi}\sigma_\epsilon^2 \cdot 2\pi\theta \\ &= \theta\sigma_\epsilon^2 \end{aligned}$$

which is the same as what we got from working in the time domain. In the computation we use the fact the $\int_{-\pi}^{\pi} e^{i\omega} d\omega = 0$, as the integral of sine or cosine functions all the way around a circle is zero.

Figure 1 plots the spectrum of MA(1) processes with positive and negative coefficients. When $\theta > 0$, we see that the spectrum is high for low frequencies and low for high frequencies. When $\theta < 0$, we observe the opposite. This is because when θ is positive, we have positive one lag correlation which makes the series smooth with only small contribution from high frequency (say, day to day) components. When θ is negative, we have negative one lag correlation, therefore the series fluctuates rapidly about its mean value.

Above we have considered the moving average process, the next proposition gives results for an ARMA models with white noise errors:

Proposition 2 *Let $\{x_t\}$ be an ARMA(p, q) process satisfying*

$$\phi(L)x_t = \theta(L)\epsilon_t$$

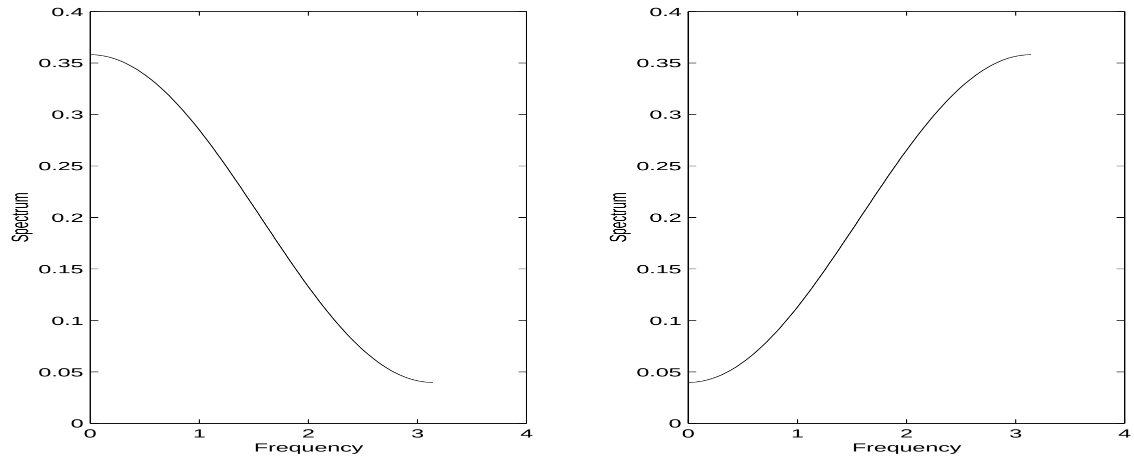

Figure 1: Plots of the spectrum of MA(1) processes ($\theta = 0.5$ for the left figure and $\theta = -0.5$ for the right figure)

where $\epsilon \sim WN(0, \sigma_\epsilon^2)$, all roots of $\phi(L)$ lies out of the unit circle, then the spectrum of x_t is:

$$\begin{aligned} S_x(\omega) &= \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} S_\epsilon(\omega) \\ &= \frac{1}{2\pi} \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} \sigma_\epsilon^2 \end{aligned}$$

Example 3 Consider an AR(1) process,

$$x_t = \phi x_{t-1} + \epsilon_t.$$

Using the above proposition,

$$\begin{aligned} S_x(\omega) &= \frac{\sigma_\epsilon^2}{2\pi} |1 - \phi e^{-i\omega}|^{-2} \\ &= \frac{\sigma_\epsilon^2}{2\pi} (1 + \phi^2 - 2\phi \cos\ \omega)^{-1} \end{aligned} \tag{8}$$

Figure 2 plots the AR(1) processes with positive and negative coefficients. We have similar observations here as the MA processes. However, note that when $\phi \to 1$, $S_x(\omega) \to \infty$, which means that a random walk process has an infinite spectrum at frequency zero. This is similar as we are working with summation and differencing. When we add up a white noise (say, $\phi = 1$ as in a random walk), the high frequencies are smoothed out (those spikes in the white noise disappear) and what is left is the long term stochastic trend. On the contrary, when we do differencing (say, do first differencing to a random walk, then we are back to the white noise series), we get rid of the long term trend, and what is left is the high frequencies (lots of spikes with mean zero, say).

Finally we introduce a *spectral representation theorem* without proof. For zero-mean stationary process with absolutely summable autocovariances, define random variables $\alpha(\omega)$ and $\delta(\omega)$, we could represent the series in the form

$$x_t = \int_0^\pi [\alpha(\omega) \cos(\omega t) + \delta(\omega) \sin(\omega t)] d\omega.$$

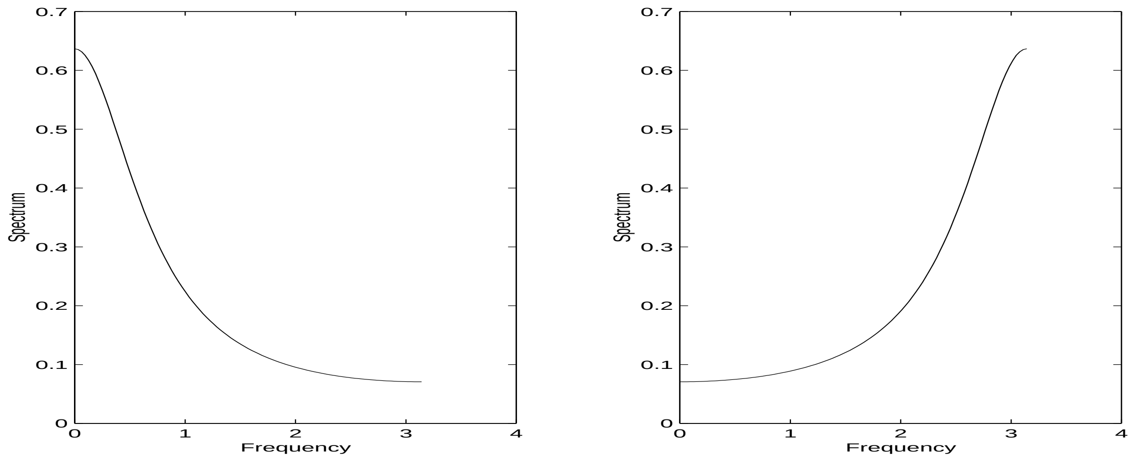

Figure 2: Plots of the spectrum of AR(1) processes ($\phi = 0.5$ for the left figure and $\phi = -0.5$ for the right figure)

where $\alpha(\omega)$ and $\delta(\omega)$ have zero mean and are mutually and serially uncorrelated. The representation theorem tells that for a stationary process with absolutely summable autocovariances, we can write it as a weighted sum of periodic functions.

4 Cross Spectrum and Spectrum of a Sum

Spectrum is an autocovariance generating function and we can use it to compute the autocovariance for a stationary process. Besides computing autocovariance of a single time series, the spectrum function can also capture the covariance cross two time series. We call such spectrum functions *cross spectrum*.

For a single time series $\{x_t\}$, a spectrum function is the Fourier transform of the autocovariance function $\gamma_x(h) = E(x_t x_{t-h})$. Similarly, for two time series $\{x_t\}$ and $\{y_t\}$, the cross spectrum is the Fourier transform of the covariance function of x_t and y_{t-h}, i.e.

$$S_{xy}(\omega) = \sum_{h=-\infty}^{\infty} e^{-ih\omega} E(x_t y_{t-h})$$

In general,

$$S_{xy}(\omega) \neq S_{yx}(\omega) = \sum_{h=-\infty}^{\infty} e^{-ih\omega} E(y_t x_{t-h})$$

But they have the following relationship:

$$S_{xy}(\omega) = \overline{S_{yx}(\omega)} = S_{yx}(-\omega)$$

which is easy to verify:

$$S_{xy}(\omega) \quad = \quad \sum_{h=-\infty}^{\infty} e^{-ih\omega} E(x_t y_{t-h})$$

$$
\begin{aligned}
&= \sum_{h=-\infty}^{\infty} e^{-ih\omega} E(y_t x_{t+h}) \\
&= \sum_{k=-\infty}^{\infty} e^{ik\omega} E(y_t x_{t-k}) \quad (\text{let} \quad k = -h) \\
&= \sum_{k=-\infty}^{\infty} e^{-(-ik\omega)} E(y_t x_{t-k}) \\
&= S_{yx}(-\omega)
\end{aligned}
$$

Note that if x_t and y_s are uncorrelated for all t, s, then $E(x_t y_{t-h}) = 0$ for all h, therefore, $S_{xy}(\omega) = S_{yx}(\omega) = 0$. Knowing the cross spectrum, next we can compute the spectrum of a sum. For a process $z_t = x_t + y_t$, the spectrum of z_t can be computed as follows:

$$
\begin{aligned}
S_z(\omega) &= \sum_{h=-\infty}^{\infty} e^{-ih\omega} E(z_t z_{t-h}) \\
&= \sum_{h=-\infty}^{\infty} e^{-ih\omega} E[(x_t + y_t)(x_{t-h} + y_{t-h})] \\
&= \sum_{h=-\infty}^{\infty} e^{-ih\omega} [E(x_t x_{t-h}) + E(x_t y_{t-h}) + E(y_t x_{t-h}) + E(y_t y_{t-h})] \\
&= S_x(\omega) + S_{xy}(\omega) + S_{yx}(\omega) + S_y(\omega)
\end{aligned}
$$

We have proposed before that for a time series z_t, its spectrum decompose its variation to different components contributed from each frequency ω. Here, we see another form of decomposition: we can decompose the variation in z to different sources. In particular, if x_t and y_s are uncorrelated for all t, s, i.e., $S_{xy}(\omega) = S_{yx}(\omega) = 0$, then we have

$$S_z(\omega) = S_x(\omega) + S_y(\omega).$$

5 Estimation

In equation (3), we define the spectrum as

$$S_x(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\omega} \gamma_x(h).$$

Given a stationary process, the sample autocovariance can be estimated

$$\hat{\gamma}_x(h) = T^{-1} \sum_{t=h+1}^{T} [(x_t - \bar{x})(x_{t-h} - \bar{x})].$$

To estimate the spectrum, we may compute the sample analog of (3), which is known as the *sample periodogram*

$$I_x(\omega) = \frac{1}{2\pi} \sum_{h=-T+1}^{T-1} e^{-ih\omega} \hat{\gamma}_x(h).$$

Or we can equivalently write it as

$$I_x(\omega) = \frac{1}{2\pi}\left(\hat{\gamma}(0) + 2\sum_{h=1}^{T-1}\hat{\gamma}(h)\cos(\omega h)\right). \tag{9}$$

We have the following asymptotic distribution of the sample periodogram.

$$\frac{2I_x(\omega)}{S_x(\omega)} \sim \chi^2(2)$$

Since $E(\chi^2(2)) = 2$, the sample periodogram provides an unbiased estimate of the spectrum, $\lim_{T\to\infty} EI_x(\omega) = S_x(\omega)$. However, the variance of $I_x(\omega)$ does not go to zero. In fact,

$$Var(I_x(\omega)) \to \begin{cases} 2S_x^2(0) & \text{for} \quad \omega = 0 \\ S^2(\omega) & \text{for} \quad \omega \neq 0 \end{cases}$$

Therefore, even when the sample size is very large, the sample periodogram still could not provide an accurate estimate for the true spectrum. To estimate the spectrum, there are two better approaches. First is parametric approach. We can estimate the ARMA model using least square or MLE to obtain consistent estimator of the parameters, and then plug in the estimator to obtain a consistent estimator for the spectrum. For instance, for an MA(1) process,

$$x_t = \epsilon_t + \theta\epsilon_{t-1}, \quad \epsilon_t \sim WN(0, 1)$$

if we could obtain a consistent estimator for θ, denoted by $\hat{\theta}$, then for any ω,

$$\hat{S}_x(\omega) = \frac{1}{2\pi}[1 + \hat{\theta}^2 + \hat{\theta}(e^{-i\omega} + e^{i\omega})].$$

A potential problem with parametric estimation is that we have to specify a parametric model for the process, say, ARMA(p, q). So we may have some errors due to misspecification. However, even if the model is incorrectly specified, if the autocovariances of the true process are close to those for our specifications, then this procedure still could provide a useful estimate of the population spectrum.

An alternative approach is to estimate the spectrum nonparametrically. Doing this could save us from specifying a model for the process. We still make use of the sample periodogram, however, to estimate the spectrum $S_x(\omega)$, we use a weighted average of the sample periodogram over several neighboring ωs. How much weight to put on each ω in the neighborhood is determined by a function which is known as the kernel, or kernel function. This means that the spectrum is estimated by

$$\hat{S}_x(\omega_j) = \sum_{l=-m}^{m} k(l, m) \cdot I_x(\omega_{j+l}). \tag{10}$$

The kernel function $k(l, m)$ must satisfy that

$$\sum_{l=-m}^{m} k(l, m) = 1.$$

Here m is the bandwidth or window indicating how many different frequencies to viewed as useful in estimating $S_x(\omega_j)$. Averaging $I_x(\omega)$ over different frequencies can equivalently be represented as multiplying the hth autocovariances $\gamma(h)$ in (9) by a weight function $w(h, q)$. A derivation can be found on page 166 on Hamilton.

These weight function $w(h, q)$ satisfy that $w(0, q) = 1$, $|w(h, q)| \leq 1$, and $w(h, q) = 0$ for $h > q$. The q in weight function works in a similar way as the m in $k(l, m)$, as it specifies a length of the window. Some commonly used weight functions are

Truncated kernel, let $x = h/q$,

$$w(x) = \begin{cases} 1 & \text{for} \quad |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Bartlett kernel, let $x = h/q$,

$$w(x) = \begin{cases} 1 - |x| & \text{for} \quad |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Modified Bartlett kernel

$$w(h, q) = \begin{cases} 1 - \frac{h}{q+1} & \text{for} \quad h = 1, 2, \ldots, q \\ 0 & \text{otherwise} \end{cases}$$

Parzen kernel, let $x = h/q$,

$$w(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3 & \text{for} \quad |x| < 1/2 \\ 2(1 - |x|)^3 & \text{for} \quad 1/2 \leq |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

A typical problem in nonparametric estimation is the trade off between variance and bias. Usually a large bandwidth reduces variance but induces bias. To reduce the variance without adding much bias, we need to choose a proper bandwidth. In practice, we may plot an estimate of the spectrum using several different bandwidths and use subjective judgment to choose the bandwidth. Basically, if the plot is too flat, then it is hard to draw information like which frequencies are more important than others; on the other hand, if the plot is too choppy (too many peaks and valleys mixed together), then it is hard to make convincing comments.

Example 4 (Spectrum estimation of an AR(1) process). The data are generated from

$$x_t = \phi x_{t-1} + \epsilon_t, \quad \phi = 0.5, \epsilon_t \sim i.i.d.N(0, 1).$$

We simulated a sequence of length $n = 200$ using this DGP and the OLS estimates of ϕ is 0.59 (OLS estimate is consistent in this problem). The upper-left figure in Figure 3 plots the population spectrum, i.e., using (8) with $\phi = 0.5$. The upper-right figure plots the estimated spectrum using (8) with the OLS estimates of ϕ, 0.59. The lower-left figure plots the sample periodogram $I_x(\omega)$, which is very volatile. Finally, the lower right figure plots the smoothed estimate for the spectrum using the Bartlett kernel, i.e.

$$\hat{S}_x(\omega) = (2\pi)^{-1} \left[\hat{\gamma}_x(0) + 2 \sum_{j=1}^{q} \left(1 - \frac{j}{q+1} \right) \hat{\gamma}_x(j) cos(\omega j) \right],$$

where q is set to be 5.

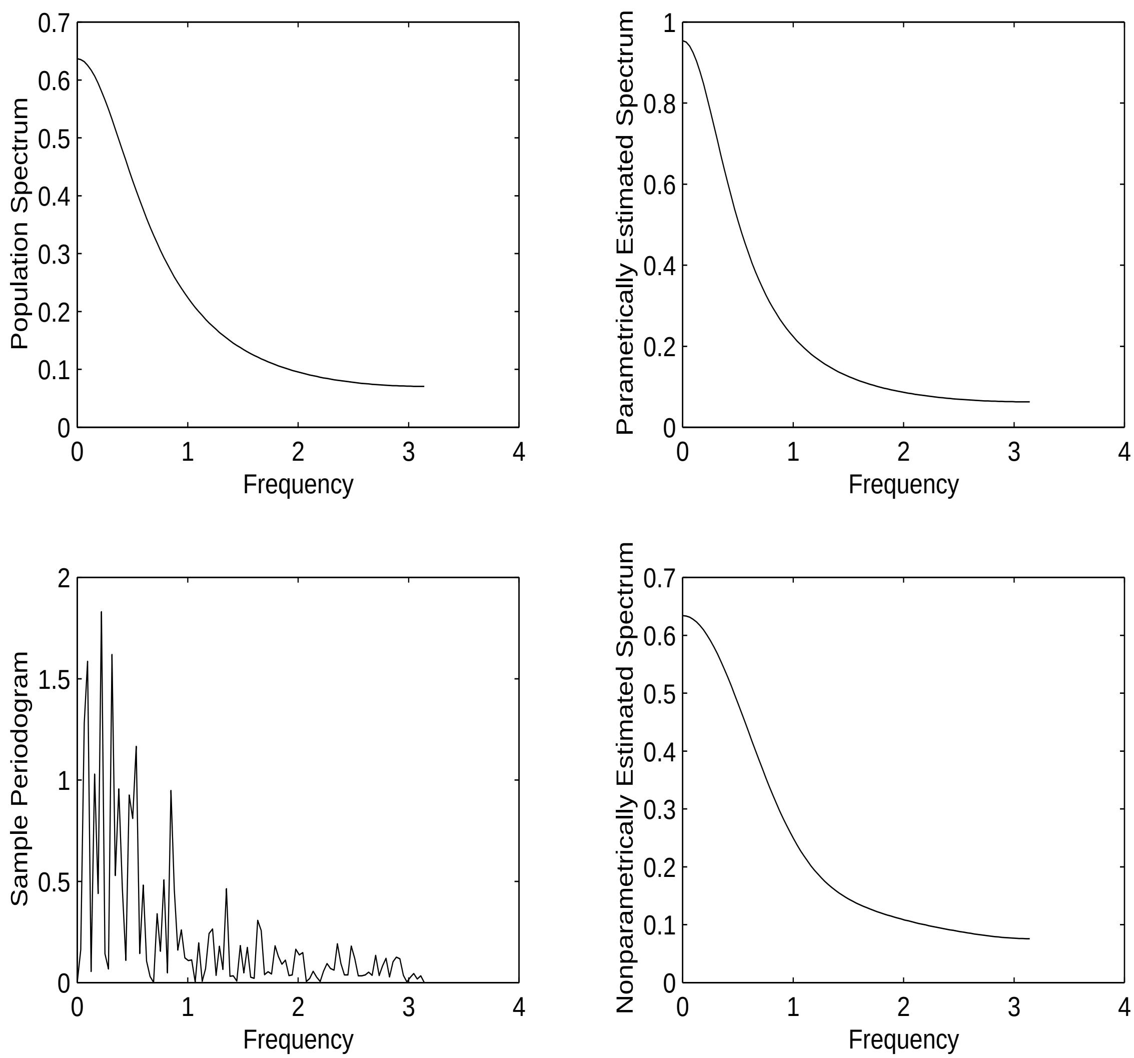

Figure 3: Estimates for Spectrum

In empirical studies, Section 6.4 on spectrum of industrial production series in Hamilton provides a very good example. Without any detrending, the spectrum is focused on the low frequency region, which means that the variance of the series is largely from the long term trend (here is the economics growth). After detrending, we obtain the growth rate which is stationary, and the variance now mostly come from the business cycle and seasonal effects. After filtering the seasonal effects, most of the variance is now due to the business cycle.

Readings: Hamilton, Ch. 6; Brockwell and Davis, Ch. 4, Ch. 10