Spectral Analysis

Spectral analysis of a stationary time series involves a change of variables so that the original autocorrelated (but homoskedastic) process is mapped into an uncorrelated (but heteroskedastic) process. The same change of variables maps the autocorrelation function of the original process into the variance function of the new process. Since the change of variables is in terms of trigonometric functions, spectral analysis is also called *frequency domain analysis* or *harmonic analysis.* The change of variables uses the Fourier transform of applied mathematics (see section 3.2 below) and hence the subject is also referred to as the *Fourier analysis* of time series. Finally, when applied to an actual finite time-series data set, the study is sometimes referred to as *periodogram* analysis. These phrases are used interchangeably in the literature.

As discussed in section 3.1 below, sine and cosine functions of a real variable x can be written in terms of the complex exponential function e^{ix}, where $i = \sqrt{-1}$. Thus spectral theory can alternatively be developed using trigonometric expressions or using complex exponential expressions. Because the latter are somewhat less messy, the complex exponential version will be emphasized in these notes.

In a sense, there are two distinct theories to be developed. *Population spectral theory* (covered in section 1) studies an infinite time series $\{y_t\}$ where t ranges over all integers from $-\infty$ to $+\infty$. The Fourier transform maps this infinite sequence of random variables into an (uncountably infinite) continuous state random process. Likewise, the infinite autocovariance sequence $\{\gamma_r\}$ is mapped into an (almost everywhere) continuous function. In this theory all parameters are assumed known; there is no estimation involved. Various ways of describing the interesting features of a population process are investigated. *Finite sample spectral theory* (covered in section 2) studies an observed sample of finite length T drawn from this population. The finite Fourier transform maps the T-dimensional sample vector into another vector of the same dimension. The analysis is closely related to the problem of diagonalizing a covariance matrix to get rid of autocorrelation. The purpose of the finite sample theory is to develop useful feasible transformations that simplify data analysis for estimation and testing of unknown parameters. Of course, as the sample size tends to infinity, the finite-sample theory begins to look very much like the population theory.

1   Population Spectral Theory

1.1   The spectral representation of an autocovariance function

Suppose $\{y_t\}$ is a mean-zero stationary process defined over the intergers and $\{\gamma_r\}$ is the corresponding autocovariance sequence. Then $\{\gamma_r\}$ must be symmetric and positive semidefinite. That is, $\gamma_r = \gamma_{-r}$ and

$$\sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \gamma_{r-s}\alpha_r \alpha_s \geq 0$$

for every sequence of real numbers $(\alpha_1, \alpha_2, ...)$. A key result is:

A necessary and sufficient condition for $\{\gamma_r\}$ to be a valid autocovariance sequence is that γ_r can be expressed as a Stieltjes integral of the form

$$\gamma_r = \int_{-\pi}^{\pi} e^{ir\lambda} dS(\lambda) = \int_{-\pi}^{\pi} \cos(r\lambda) dS(\lambda) \tag{1}$$

where $S(\lambda)$ is a monotonically nondecreasing function defined on $[-\pi,\pi]$ and $dS(\lambda)=dS(-\lambda)$.

Sketch of proof: If $\gamma_r = \int_{-\pi}^{\pi} e^{ir\lambda} dS(\lambda)$, then

$$\sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \gamma_{r-s} \alpha_r \alpha_s = \int_{-\pi}^{\pi} \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \alpha_r \alpha_s e^{i(r-s)\lambda} dS(\lambda) = \int_{-\pi}^{\pi} |\sum_{r=1}^{\infty} \alpha_r e^{ir\lambda}|^2 dS(\lambda)$$

which is nonnegative if $S(\lambda)$ is a nondecreasing function. Conversely, define

$$
\begin{aligned}
f_N(\lambda) &= \frac{1}{2\pi N} E \left| \sum_{t=1}^{N} y_t e^{i\lambda t} \right|^2 = \frac{1}{2\pi N} \sum_{t=1}^{N} \sum_{s=1}^{N} \gamma_{s-t} e^{-i\lambda(s-t)} \\
&= \frac{1}{2\pi} \sum_{k=-N+1}^{N-1} \gamma_k (1 - \frac{|k|}{N}) e^{-i\lambda k} \equiv \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} b_N(k) e^{-i\lambda k}
\end{aligned}
$$

where $b_N(k)$ is zero when $k \geq N$. Clearly $f_N(\lambda) \geq 0$ and the sequence $b_N(1), b_N(2), ...$ is absolutely summable for fixed N. A basic result of Fourier theory states that if $\{c_k\}$ is a doubly infinite absolutely summable symmetric sequence then the infinite sum $\frac{1}{2\pi} \sum_k c_k e^{-i\lambda k}$ converges uniformly on $[-\pi, \pi]$ to a continuous even function $f(\lambda)$ and that $c_k = \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda k} d\lambda$. Thus $f_N(\lambda)$ is a nonnegative continuous even function and, for $k < N$,

$$b_N(k) = \gamma_k (1 - \frac{|k|}{N}) = \int_{-\pi}^{\pi} e^{ik\lambda} f_N(\lambda) d\lambda = \int_{-\pi}^{\pi} e^{ik\lambda} dF_N(\lambda)$$

where $F_N(\lambda) = \int_{-\pi}^{\lambda} f_N(x) dx$ is a bounded monotonically nondecreasing function. By Helly's theorem, there is a subsequence along which F_N converges to F. As N tends to infinity along this sequence we find $\gamma_r = \int_{-\pi}^{\pi} e^{ir\lambda} dF(\lambda)$ where $F(\lambda)$ is monotonically nondecreasing.

The function $S(\lambda)$ is called the *spectral distribution function* for the sequence $\{\gamma_r\}$. By integrating $f_N(\lambda)$ and letting N tend to infinity we can find an explicit expression (up to an arbitrary constant) for $S(\lambda)$. If normalization is accomplished by setting $S(0) = 0$, we find

$$S(\lambda) = \frac{1}{2\pi} [\gamma_0 \lambda + 2 \sum_{r=1}^{\infty} \gamma_r \frac{\sin \lambda r}{r}] \tag{2}$$

as long as we follow the convention that, at points of discontinuity, $S(\lambda)$ is defined as $[S(\lambda+) + S(\lambda-)]/2$. Typically (2) is not especially useful for actually computing $S(\lambda)$. However, we do easily obtain the results $S(-\pi) = -\gamma_0/2$ and $S(\pi) = -\gamma_0/2$.

If $\sum_r |\gamma_r| < \infty$, $S(\lambda)$ is everywhere differentiable with continuous derivative $s(\lambda)$ and equations (1) and (2) simplify to the usual Fourier transform pair

$$\gamma_r = \int_{-\pi}^{\pi} e^{ir\lambda} s(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(\lambda r) \, s(\lambda) d\lambda \tag{1'}$$

$$s(\lambda) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \gamma_r e^{-ir\lambda} = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \gamma_r \cos(r\lambda). \tag{2'}$$

The derivative function $s(\lambda)$ is called the *spectral density function* for the sequence $\{\gamma_r\}$ (or *spectrum* for short.) By construction $s(\lambda)$ is continuous, nonnegative, and symmetric about zero. Note that $2\pi s(\lambda) = G(e^{-i\lambda})$ where $G(L)$ is the autocovariance generating function. Equation (1') implies that $\int s(\lambda) d\lambda = \gamma_0$; that is, the area under the spectral density function is the variance. Equation (2') implies that the spectral density at zero is $(2\pi)^{-1} \sum_r \gamma_r$.

If the autocovariances are square summable, but not absolutely summable, the results in the previous paragraph have to be slightly modified. Equation (2') will no longer necessarily

be an everywhere convergent series defining a continuous function. But there does exist a function $s(\lambda)$ such that

$$\int_{-\pi}^{\pi} |s(\lambda) - \frac{1}{2\pi}\sum \gamma_r e^{-ir\lambda}|^2 d\lambda = 0 \ .$$

That is, we have convergence in the mean. The function $s(\lambda)$ may be unbounded at one or more isolated points, but it is square integrable and (1') holds.

Recall that Wold's representation theorem says that every weakly stationary process can be written as the sum of two uncorrelated processes: a moving average process with square-summable coefficients and a perfectly predictable deterministic process. Corresponding to this decomposition, the spectral distribution function $S(\lambda)$ can be written as the sum of two functions: one that is differentiable and the other is typically a step function. For the first function, (1') and (2') are valid, although in the case where the moving average coefficients are not absolutely summable $s(\lambda)$ may tend to infinity at isolated points. (An example is given in section 1.3.) An example of a deterministic process with step-function for $S(\lambda)$ is the so-called harmonic process

$$y_t = \sum_{j=1}^{p} [U_j \sin(\theta_j t) + V_j \cos(\theta_j t)]$$

where the U's and V's are uncorrelated random variables with zero means and finite variances. Based on any 2p observations on y_t one can calculate the U's and the V's. All future observations are then perfectly predictable. The spectral distribution function for this process is a step function and $dS(\lambda)$ behaves like a mass function with spikes at the values $\pm\theta_1, ..., \pm\theta_p$; equation (1') remains valid if the integrals are replaced by a sum. For most of these notes we shall restrict ourselves to the case where $\{y_t\}$ has no deterministic component and $S(\lambda)$ is continuously differentiable except possibly at zero.

1.2 The spectral representation of a stationary process

Suppose $\{y_t\}$ is a weakly stationary mean-zero discrete-time stochastic process whose auto-covariance sequence has spectral distribution function $S(\lambda)$. Fourier theory may be applied to this random sequence as well. The basic result (known as the *Cramér representation theorem*) is that y_t can always be written as a stochastic integral

$$y_t = \int_{-\pi}^{\pi} e^{i\lambda t} dH(\lambda) \tag{3}$$

where $H(\lambda)$ is a continuous-time complex-valued stochastic process defined on $[-\pi, \pi]$. The process $H(\lambda)$ has uncorrelated increments, mean zero, and $E|dH|^2 = dS$. If $S(\lambda)$ is differentiable, (3) can be written as

$$y_t = \int_{-\pi}^{\pi} e^{i\lambda t} \sqrt{s(\lambda)} dW(\lambda)$$

where $W(\lambda)$ is a complex-valued process with uncorrelated increments, mean zero, and $E|W(\lambda)|^2 = \lambda$. When the data are normal, $W(\lambda)$ is a complex Brownian motion on $[-\pi, \pi]$. The inverse Fourier transform can be written, up to a random additive constant, as

$$H(\lambda) = \frac{1}{2\pi}[y_0\lambda + 2\sum_{t\neq 0} y_t \frac{e^{-i\lambda t}}{-it}]$$

where again, at points of discontinuity, $H(\lambda)$ is defined as $[H(\lambda+) + H(\lambda-)]/2$. Athough the inverse expression has no simple interpretation, the Cramér representation (3) can be rewritten is a way that sheds considerable light on the meaning of the spectrum.

Using the fact that y_t is real and that $e^{i\lambda t} = \cos(\lambda t) + i\sin(\lambda t)$, we can rewrite equation (3) as

$$y_t = \int\limits_0^\pi [\cos(\lambda t)dA(\lambda) + \sin(\lambda t)dB(\lambda)] \qquad (3')$$

where $A(\lambda)$ and $B(\lambda)$ are uncorrelated continuous-time real-valued stochastic processes defined on $[0, \pi]$. The random functions have mean zero and uncorrelated increments. If we normalize so $A(0) = B(0) = 0$, then $E|A(\lambda)|^2 = E|B(\lambda)|^2 = 2S(\lambda)$. We interpret equation (3') as saying that every discrete-time stationary process can be viewed as an (uncountable) linear combination of sine and cosine functions with uncorrelated random weights. For $0 \le a < b \le \pi$, $2[S(b) - S(a)]$ is the fraction of the total variance of y_t attributable to the sine and cosine functions with frequencies between a and b.

When $S(\lambda)$ is differentiable so the spectral density $s(\lambda)$ exists, the stochastic integral (3') can be rewritten as

$$y_t = \int\limits_0^\pi [\cos(\lambda t)\sqrt{2s(\lambda)}dW_1(\lambda) + \sin(\lambda t)\sqrt{2s(\lambda)}dW_2(\lambda)]$$

where $W_1(\lambda)$ and $W_2(\lambda)$ are uncorrelated continuous-time zero-mean stochastic processes defined on $[0, \pi]$ with $E|W_1(\lambda)|^2 = E|W_2(\lambda)|^2 = \lambda$. Then $2s(\lambda)d\lambda$ is the fraction of the total variance of y_t attributable to the sine and cosine functions with frequencies in the interval $(\lambda, \lambda + d\lambda)$. If the data are normal, $W_1(\lambda)$ and $W_2(\lambda)$ are independent Brownian motions.

1.3 Some reasons for studying the population spectrum

The autocovariance function and the spectrum are Fourier transform pairs; all the information in one is summarized in (and can be recovered from) the other. The spectrum is often more convenient to work with because of the following property: Suppose $y_t = C(L)x_t + u_t$, where $\{x_t\}$ and $\{u_t\}$ are independent mean-zero stationary processes with spectral densities $s_x(\lambda)$ and $s_u(\lambda)$. The autocovariance function for $\{y_t\}$ is a complicated function of the autocovariance functions of $\{x_t\}$ and $\{u_t\}$. But, if the lag coefficients are square summable, $s_y(\lambda)$ exists and is given by

$$s_y(\lambda) = s_x(\lambda)C(e^{-i\lambda})C(e^{i\lambda}) + s_u(\lambda) = s_x(\lambda)|C(e^{-i\lambda})|^2 + s_u(\lambda) \qquad (4)$$

where $C(e^{-i\omega}) = \sum_{-\infty}^{\infty} c_r e^{-i\omega r}$ is the Fourier transform of the sequence $\{c_r\}$. Thus, the spectrum is a natural tool for studying the effect of a filter on a time series. (A derivation of (3) appears in Section 3.4 below. Note that $C(e^{-i\omega})$ will generally be complex unless $C(L)$ is a two-sided filter with $c_r = c_{-r}$ for all r.)

Since the spectrum of white noise is $\sigma^2/2\pi$, it follows from equation (4) that the spectrum of the invertible ARMA process $A(L)y_t = B(L)\varepsilon_t$ is

$$s_y(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|B(e^{-i\lambda})|^2}{|A(e^{-i\lambda})|^2} \ .$$

For example, the spectrum of the AR(1) process $y_t = \alpha y_{t-1} + \varepsilon_t$ when $|\alpha| < 1$ is

$$s_y(\lambda) = \frac{1}{2\pi}\frac{\sigma_\varepsilon^2}{(1 - \alpha e^{-i\lambda})(1 - \alpha e^{i\lambda})} = \frac{1}{2\pi}\frac{\sigma_\varepsilon^2}{1 + \alpha^2 - 2\alpha\cos\lambda}.$$

The spectrum of the fractionally differenced process $(1 - L)^d y_t = \varepsilon_t$ is

$$s_y(\lambda) = \frac{1}{2\pi}\sigma_\varepsilon^2|1 - e^{i\lambda}|^{-2d} = \frac{1}{2\pi}\sigma_\varepsilon^2[4\sin^2(\lambda/2)]^{-d}$$

which (since $\sin x \approx x$ when x is near zero) is approximately proportional to λ^{-2d} near the origin.

Some additional advantages of the spectral approach are:

4

1. Not every sequence $\{\gamma_r\}$ can be a valid autocovariance sequence; the constraint that the sequence be positive semi-definite is hard to verify. But we have the simple result: the absolutely summable sequence $\{\gamma_r\}$ is a valid autocovariance sequence if and only if $\sum_r \gamma_r e^{-i\lambda r}$ is non-negative on $[-\pi, \pi]$.

2. Cyclical behavior is more naturally studied using the spectrum rather than the autocorrelation function. A process whose spectrum is fairly flat except for a pronounced peak at frequency λ_0 behaves a lot like a random sine wave with frequency λ_0. In particular, seasonal patterns and business cycles are conveniently investigated using the spectrum. For example, one can try to design optimal filters to remove seasonality from a time series.

3. When studying vector processes, leads and lags among variables have simple representations in the frequency domain.

4. Since spectral analysis can be interpreted as a change in coordinate system such that a stationary autocorrelated series is transformed into a heteroskedastic series, results previously derived for uncorrelated but heteroskedastic models can be used in the study of stationary time series models. This point is explored in section 2.4 below where we examine how the Gaussian likelihood function can be approximated using the sample spectrum.

5. Inference on the spectrum is somewhat easier than inference on the autocorrelation function. Two autocovariance estimates, say $\widehat{\gamma}_j$ and $\widehat{\gamma}_k$, are typically highly correlated; and their covariance matrix depends in general on the entire autocovariance function. Two spectral estimates, say $\widehat{s}(\lambda_1)$ and $\widehat{s}(\lambda_2)$, are approximately uncorrelated; the variance of $\widehat{s}(\lambda)$ depends approximately only on $s(\lambda)$. Again, this point is elaborated in sections 2.1 and 2.2 below.

2   Finite-sample Spectral Theory

2.1   The finite Fourier transform

Now our attention will be on the Fourier transform associated with a finite time series. To motivate its use consider the following algebraic fact: If $\mathbf{y}$ is a T-dimensional random column vector with mean zero and covariance matrix $\mathbf{V}$, then there exists an orthogonal matrix $\mathbf{Q}$ such that

$$\mathbf{D} = \mathbf{Q}'\mathbf{V}\mathbf{Q} \tag{5}$$

is diagonal with nonnegative elements. (Since $\mathbf{V}$ is positive semidefinite, such a $\mathbf{D}$ and $\mathbf{Q}$ can always be constructed from the characteristic roots and vectors of $\mathbf{V}$.) Then it follows that

1. The transformed variable $\widetilde{\mathbf{y}} = \mathbf{Q}'\mathbf{y}$ has uncorrelated elements and the quadratic form $\mathbf{y}'\mathbf{V}^{-1}\mathbf{y}$ that appears in the normal likelihood function simplifies to a sum of squares $\widetilde{\mathbf{y}}'\mathbf{D}^{-1}\widetilde{\mathbf{y}}$.

2. The vector $\mathbf{y}$ can be written as $\mathbf{y} = \mathbf{Q}\mathbf{D}^{1/2}\varepsilon$ where ε is white noise with unit variance. If $\mathbf{q}_j$ is the jth column of $\mathbf{Q}$ and d_j is the jth diagonal element of $\mathbf{D}$, then we can write

$$\mathbf{y} = \mathbf{q}_1\sqrt{d_1}\varepsilon_1 + \cdots + \mathbf{q}_T\sqrt{d_T}\varepsilon_T. \tag{6}$$

That is, $\mathbf{y}$ can be viewed as a linear combination of T deterministic series with uncorrelated random weights; d_j is the variance of the weight given to series j.

Unfortunately, these facts are generally not very useful. The matrix $\mathbf{Q}$ typically depends on the unknown elements of $\mathbf{V}$ so $\widetilde{\mathbf{y}}$ is not observable; and computing $\mathbf{Q}$ is just as hard as computing $\mathbf{V}^{-1}$. Furthermore, the characteristic vectors $\mathbf{q}_j$ are typically uninterpretable so the decomposition (6) is uninformative. The special case where $\mathbf{y}$ is the realization of a stationary time series is an exception. In that case, when T is large, $\mathbf{V}$ is approximately diagonalized by a *known* matrix and the decomposition is interpretable in terms of *latent cycles*. These results follow from the finite version of the population Fourier theory discussed in Section 1.

Suppose T is an odd number so $T = 2m + 1$ for some integer $m > 0$. We define the *Fourier frequencies*

$$\lambda_k = 2\pi \frac{k}{T} \qquad \text{for } k = 0, \pm 1, \pm 2, \cdots, \pm m$$

These numbers divide the interval $[-\pi, \pi]$ into $T + 1$ subintervals, all but the outer two of equal length. (For even T, $k = 0, \pm 1, \pm 2, \cdots, \pm (T-2)/2, T/2$; the following analysis is only slightly changed.) Let $\mathbf{F}$ be the $T \times T$ *Fourier matrix* whose tk element is given by

$$f_{tk} = \frac{1}{\sqrt{T}} \exp\{it\lambda_k\} = \frac{1}{\sqrt{T}}[\cos(t\lambda_k) + i\sin(t\lambda_k)]. \tag{7}$$

The key result is: $\mathbf{F}$ is a (complex) orthogonal matrix and it almost diagonalizes the covariance matrix of any stationary process with continuous spectrum. That is $\mathbf{F}^*\mathbf{F} = I$ and $\mathbf{F}^*\mathbf{VF}$ is almost diagonal if T is large and $\mathbf{V}$ has the Toeplitz form with ij element given by $\gamma_{|i-j|}$ and $\sum_r |\gamma_r| < \infty$. (By $*$ we mean "transpose and take complex conjugate.") Note that the diagonal elements of $\mathbf{F}^*\mathbf{VF}$ are linear functions of the autocovariances $(\gamma_0, \gamma_1, ..., \gamma_{T-1})$. Indeed , the kth diagonal element is approximately

$$\sum_{r=-\infty}^{\infty} \gamma_r e^{-i\lambda_k r} = 2\pi s(\lambda_k). \tag{8}$$

Sketch of proof. Write the pq element of F^*F as

$$\frac{1}{T}\sum_{t=1}^{T} e^{-i\lambda_p t} e^{i\lambda_q t} = \frac{1}{T}\sum_{t=1}^{T} e^{2\pi i(q-p)t/T}$$

which is one if $p = q$ and zero otherwise. (See Section 3.4 below for a proof.) Write the pq element of F^*VF as

$$
\begin{aligned}
a_{pq} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} e^{-i\lambda_p t} e^{i\lambda_q s} v_{ts} \\
&= \frac{1}{T}\left[\gamma_0 R_T(0) + \sum_{r=1}^{T-1} \gamma_r e^{-i\lambda_p r} R_T(r) + \sum_{r=-1}^{-T+1} \gamma_r e^{-i\lambda_q r} R_T(r)\right]
\end{aligned}
$$

where

$$R_T(r) = \sum_{t=1}^{T-|r|} e^{i(\lambda_q - \lambda_p)t}.$$

When $p = q$, $R_T(r) = T - |r|$ and a_{pp} is given by

$$\sum_{r=-T+1}^{T-1} \gamma_r e^{-i\lambda_p r}\left(1 - \frac{|r|}{T}\right). \tag{9}$$

When $p \neq q$, it follows from Section 3.4 that $|R_T(r)| < |r|$. As long as $|r\gamma_r| \to 0$ as $r \to \infty$, it follows that $|a_{pq}| \to 0$ as $T \to \infty$. Using a more tedious argument, the condition $|r\gamma_r| \to 0$ can be relaxed.

This result can be interpreted as saying that, for large T, the sine and cosine functions making up the columns of $\mathbf{F}$ are approximate characteristic vectors of $\mathbf{V}$ and the spectrum evaluated at the Fourier frequencies are (except for a factor of 2π) the approximate characteristic roots of $\mathbf{V}$.

2.2   Estimating the spectrum

For a given sample $\mathbf{y}$, the elements of the transformed time series $\widetilde{\mathbf{y}} = \mathbf{F}^*\mathbf{y}$ are approximately uncorrelated. From the algebra above, the variance of $\widetilde{y}_k$ is approximately $2\pi s(\lambda_k)$. Indeed,

$$
\begin{aligned}
p_k &\equiv \frac{|\widetilde{y}_k|^2}{2\pi} = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} y_t e^{-i\lambda_k t}\right|^2 = \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T} y_t y_s e^{-i\lambda_k t} e^{i\lambda_k s} \\
&= \frac{1}{2\pi T}\sum_{r=-T+1}^{T-1} e^{-i\lambda_k r}\sum_{t=1}^{T-|r|} y_t y_{t+r} = \frac{1}{2\pi}\sum_{r=-T+1}^{T-1} \widehat{\gamma}_r e^{-i\lambda_k r}
\end{aligned}
$$

as long as $\mathrm{E}y_t = 0$. (If $\mathrm{E}y_t \neq 0$, $y_t - \bar{y}$ should be used in place of y_t.) Hence, p_k is the natural estimate of the spectrum at frequency λ_k. Although approximately unbiased, it is not a consistent estimate since its variance does not go to zero as $T \to \infty$. Indeed, since

$$
\tilde{y}_k = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_t e^{-i\lambda_k t} = \sqrt{T}\left[\frac{1}{T}\sum_{t=1}^{T} y_t \cos(\lambda_k t) + \frac{i}{T}\sum_{t=1}^{T} y_t \sin(\lambda_k t)\right] \equiv \sqrt{T}[U_T + iV_T]
$$

where both U_T and V_T are weighted sample averages of zero-mean weakly dependent random variables, we would expect $\sqrt{T}U_T$ and $\sqrt{T}V_T$ to be asymptotically normal. When $k \neq 0$, they turn out to be asymptotically uncorrelated with variance $\pi s(\lambda_k)$ so $2p_k/s(\lambda_k)$ is approximately distributed as $\chi^2(2)$; when $k = 0$, $V_T = 0$ and $2p_0/s(\lambda_0)$ is approximately distributed as $\chi^2(1)$. The limiting distribution of p_k is a random variable not a constant so it is not a consistent estimator.

As long as the true spectrum is a smooth function of λ, a consistent estimate of the spectrum can be obtained by fitting nonparametrically a smooth curve through the points (p_k, λ_k). The graph of p_k against λ_k is often called the *periodogram* and the p_k are called *periodogram values*. (Note that $p_k = p_{-k}$, so we need only look at the nonnegative frequencies where $k = 0, ..., m$.)

A smoothed periodogram estimate of the spectrum at frequency λ typically looks like

$$
\widehat{s}(\lambda) = \sum_{k=-m}^{m} K_T(\lambda_k - \lambda)p_k
$$

where $K_T(\omega)$ is a continuous weighting function (called a kernel) that decreases in magnitude as $|\omega|$ moves away from zero. By averaging over the (approximately) uncorrelated periodogram values, variance is reduced but at the cost of introducing bias. The kernel will be chosen so that, for any $\omega \neq 0$, $K_T(\omega) \to 0$ as $T \to \infty$. If the rate of convergence is cleverly chosen, it can be shown that $\widehat{s}(\lambda)$ is a consistent estimate of $s(\lambda)$; that is, both the bias and the variance go to zero as $T \to \infty$. Using the definition of p_k, the smoothed spectral estimate can always be rewritten as

$$
\widehat{s}(\lambda) = \sum_{r=-T+1}^{T-1} d_{r,T}\widehat{\gamma}_r e^{-i\lambda r}
$$

where the d's are the Fourier coefficients of the function $K_T(\omega)$. For reasonable kernels, the poorly estimated high-order autocovariances are downweighted.

2.3 Some more asymptotics

As T tends to infinity, the λ_k fill up the entire interval $[-\pi, \pi]$ so we can think of λ taking on any value in that interval. Thus equation (8) is, in the limit, just equation (2'). Note that the analog of equation (6) is the equation $\mathbf{y} = \mathbf{F}\widetilde{\mathbf{y}}$ which can be written in scalar notation as

$$y_t = \frac{1}{\sqrt{T}} \sum\nolimits_k e^{i\lambda_k t} \sqrt{2\pi s(\lambda_k)} \varepsilon_k$$

where the elements $\varepsilon_k = \tilde{y}_k / \sqrt{2\pi s(\lambda_k)}$ are approximately uncorrelated complex random variables with zero mean and variance one. For $0 \le \mu \le 1$, we define $[(2\mu - 1)m]$ to be $(2\mu - 1)m$ rounded down to the nearest integer. Then the standardized empirical distribution function for $\{\varepsilon_k\}$

$$W_T^*(\mu) = \frac{1}{\sqrt{T}} \sum_{k=-m}^{[(2\mu-1)m]} \varepsilon_k$$

is a step function on the unit interval with jumps equal to $\varepsilon_k / \sqrt{T}$. Under normality, $W_T^*(\mu)$ converges to Brownian motion on (0,1) and $W_T(\lambda) \equiv W_T^*(\frac{\lambda+\pi}{2\pi})/\sqrt{2\pi}$ converges to Brownian motion on the interval $(-\pi, \pi)$. Thus y_t is an approximation (based on the partition of $(-\pi, \pi)$ induced by the λ_k) to the Stieltjes integral

$$\int_{-\pi}^{\pi} e^{i\lambda t} \sqrt{s(\lambda)} dW_T(\lambda).$$

As T tends to infinity, we obtain the spectral representation found in section 1.2

$$y_t = \int_{-\pi}^{\pi} e^{i\lambda t} \sqrt{s(\lambda)} dW(\lambda).$$

2.4 Parametric estimation in the frequency domain

If the data are normal, the approximate log likelihood function for $\mathbf{y}$ is, except for an additive constant,

$$-\frac{1}{2} \log |V| - \frac{1}{2} y' V^{-1} y \approx -\frac{1}{2} \sum_{k=-m}^{m} \log s(\lambda_k) - \frac{1}{2} \sum_{k=-m}^{m} \frac{p_k}{s(\lambda_k)}.$$

This is often called the Whittle log likelihood. If one has a parametric model for the spectrum (e.g., one knows the $\{y_t\}$ are an ARMA(p,q) process), the unknown parameters could be estimated by maximizing this function.

An interesting example is linear regression with nonstochastic regressors and autocorrelated errors. Suppose $y = X\beta + u$ where the T elements of the error vector u are a stationary series with mean zero and covariance matrix V. Multiplying on the left by the transposed Fourier matrix F^*, we obtain $\tilde{y} = \tilde{X}\beta + \tilde{u}$, where the elements of $\tilde{u}$ are approximately uncorrelated. Approximate GLS estimates of β are obtained by minimizing

$$\sum |\tilde{y}_k - \beta' \tilde{x}_k|^2 / s_u(\lambda_k)$$

where s_u is the spectrum of the $\{u_t\}$ process. Feasible GLS replaces s_u by some estimate, say by smoothing the periodogram values computed from the OLS residuals or by fitting a parametric ARMA model to those residuals.

2.5   Avoiding complex numbers

The theory developed in sections 2.1-2.3 is based on the complex Fourier matrix $\mathbf{F}$. There exists an equivalent real matrix $\mathbf{Q}$ that does the same job. Since the real version is more tedious to express and it is harder to verify its orthogonality properties, I have used the complex version. In actual computations (particularly in the regression example discussed above) it is often more convenient to use the real version.

Let $\mathbf{1}$ be a T-dimensional column vector of ones. Again assume $T = 2m + 1$. For $k = 1, ..., m$, let $\mathbf{c}_k$ be the T-dimensional vector whose t'th element is $\sqrt{2/T}\cos(\lambda_k t)$ and let $\mathbf{d}_k$ be the T-dimensional vector whose t'th element is $\sqrt{2/T}\sin(\lambda_k t)$. Then $\tilde{y}_0 = \mathbf{1}'y/\sqrt{T}$ and $\tilde{y}_k = (\mathbf{c}_k - i\mathbf{d}_k)'y$ for $k = 1, ..., m$. Hence, $|\tilde{y}_0|^2 = (\mathbf{1}'y)^2/T$ and $|\tilde{y}_k|^2 = (\mathbf{y}'\mathbf{c}_k)^2 + (\mathbf{y}'\mathbf{d}_k)^2$ when $k \neq 0$.

Define the $T \times T$ matrix $\mathbf{Q} = [T^{-1/2}\mathbf{1}, \mathbf{c}_1, \mathbf{d}_1, ..., \mathbf{c}_m, \mathbf{d}_m]$. It can be shown that $\mathbf{Q}$ is a real orthogonal matrix and $\mathbf{Q}'\mathbf{V}\mathbf{Q}$ is approximately diagonal. Let $\hat{\mathbf{y}} = \mathbf{Q}'\mathbf{y}$ and write the components of $\hat{\mathbf{y}}'$ as $(a_0, a_1, b_1, a_2, b_2, ..., a_m, b_m)'$. Then $a_0 = \tilde{y}_0$ and

$$a_k = \mathbf{y}'\mathbf{c}_k \quad \text{and} \quad b_k = \mathbf{y}'\mathbf{d}_k \quad \text{for } k = 1, ..., m.$$

The periodogram value at frequency λ_k is $p_k = (a_k^2 + b_k^2)/2\pi$ for $j = 1, ..., m$ and $p_0 = a_0^2/2\pi$. If one fits the equation

$$y_t = \frac{a_0}{\sqrt{T}} + \sqrt{\frac{2}{T}} \sum_{k=1}^{m} [a_k \cos(\lambda_k t) + b_k \sin(\lambda_k t)] \qquad t = 1, ..., T$$

by least squares one obtains the regression coefficient vector $(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{y} = \mathbf{Q}'\mathbf{y} = \hat{\mathbf{y}}$. Thus the periodogram values are easily obtained from a least squares regression on sinusoidal variables. Such regressions were popular in the early 1900's under the name periodogram analysis. Now such regressions are interpreted as a way to compute spectral estimates.

3   Some Algebra

3.1   The complex exponential function

The real exponential function e^x has the key properties $e^0 = 1$ and $de^{ax}/dx = ae^{ax}$. It is natural to require the complex exponential function e^{ix} to satisfy $e^{i0} = 1$ and $de^{ix}/dx = ie^{ix}$. Let $f(x) = e^{ix} = a(x) + ib(x)$. The second condition means that $a''(x) = -a(x)$ and $b''(x) = -b(x)$. These differential equations have solutions of the form

$$
\begin{aligned}
a(x) &= A\cos(x) + B\sin(x) \\
b(x) &= C\cos(x) + D\sin(x)
\end{aligned}
$$

The conditions $f(0) = 1$ and $f'(0) = i$ require $A = D = 1$ and $B = C = 0$. Thus, we must define

$$e^{ix} = \cos(x) + i\sin(x).$$

As a bonus, we can verify that the above definition implies $e^{ix}e^{iy} = e^{i(x+y)}$, a condition we might also want to require. Finally, using the definition, we find the useful representations

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2} \quad \text{and} \quad \sin(x) = \frac{e^{ix} - e^{-ix}}{2i}$$

3.2 Elements of Fourier analysis

Fourier analysis studies the approximation of functions by trigonometric series. A basic theorem of Fourier analysis states: Let $f(x)$ be a continuous, even, nonnegative function defined on $[-\pi, \pi]$. Then $f(x)$ can be written as an absoultely convergent infinite sum

$$f(x) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} c_r e^{-irx} = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} c_r \cos(rx) \tag{10}$$

where the sequence $\{c_k\}$ is determined by

$$c_r = \int_{-\pi}^{\pi} e^{irx} f(x) dx = \int_{-\pi}^{\pi} \cos(rx) f(x) dx. \tag{11}$$

Conversly, if $\{c_k\}$ is an absolutely summable sequence with $c_k = c_{-k}$, then there exists a continuous, even, nonnegative function $f(x)$ satisfying (10) and (11). The function $f(x)$ is said to be the Fourier transform of the sequence $\{c_k\}$; the c's are called the Fourier coefficients of f. (This terminology is used even if the c's are not absolutely summable and f is not everywhere continuous. Indeed, much of Fourier analysis in concerned with extending the theory to those cases.)

It is easy to verify equation (11). If we multiply both sides of (10) by e^{ikx} and integrate, we get

$$\int_{-\pi}^{\pi} f(x) e^{ikx} dx = \int_{-\pi}^{\pi} \left[\frac{1}{2\pi} \sum_{r=-\infty}^{\infty} c_r e^{-irx} \right] e^{ikx} dx = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} c_k \left[\int_{-\pi}^{\pi} e^{i(r-k)x} dx \right] = c_r. \tag{12}$$

The last equality follows from the fact that, for integer s,

$$\int_{-\pi}^{\pi} e^{isx} dx = \int_{-\pi}^{\pi} [\cos(sx) + i\sin(sx)] dx = \begin{cases} 0 & s \neq 0 \\ 2\pi & s = 0 \end{cases}.$$

[Note: interchanging the order of summation and integration in (12) is allowed because of the absolute summability.]

3.3 Derivation of equation (4)

Let $\gamma_y(r)$ and $\gamma_u(r)$ be the autocovariance functions for the independent time series $\{y_t\}$ and $\{u_t\}$. Then, if $y_t = C(L)x_t + u_t$,

$$\gamma_y(r) = E(y_t y_{t+r}) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_j c_k \gamma_x(r + j - k) + \gamma_u(r).$$

Hence, the spectrum for $\{y_t\}$ is

$$\begin{aligned}
s_y(\lambda) &= \sum_r \sum_j \sum_k c_j c_k \gamma_x(r + j - k) e^{-i\lambda r} + \sum_r \gamma_u(r) e^{-i\lambda r} \\
&= \sum_r \sum_j \sum_k c_j e^{i\lambda j} c_k e^{-i\lambda k} \gamma_x(r + j - k) e^{-i\lambda(r+j-k)} + s_u(\lambda) \\
&= \sum_s \gamma_x(s) e^{-i\lambda s} \sum_j c_j e^{i\lambda j} \sum_k c_k e^{-i\lambda k} + s_u(\lambda) \\
&= s_x(\lambda) C(e^{i\lambda}) C(e^{-i\lambda}) + s_u(\lambda).
\end{aligned}$$

3.4   Proof that the matrix **F** is orthogonal

For complex variable z, define $S = z + z^2 + ... + z^T$. Then, since $zS = z^2 + z^3 + ... + z^{T+1}$, we obtain by subtraction

$$S = z\frac{1 - z^T}{1 - z}.$$

If $z^T = 1$ and $z \neq 1$, then $S = 0$. In particular, let $z = e^{2\pi ik/T}$ where k is an integer with $0 < |k| < T$. Then

$$
\begin{aligned}
z^T &= e^{2\pi ik} = \cos(2\pi k) + i\sin(2\pi k) = 1 \\
z &= e^{2\pi ik/T} = \cos(2\pi k/T) + i\sin(2\pi k/T) \neq 1
\end{aligned}
$$

since $\sin(2\pi k) = 0$ and $\cos(2\pi k) = 1$ for integer k and $\sin(2\pi\theta) \neq 0$ for $0 < |\theta| < 1$ except for $\theta = .5$ when $\cos(2\pi\theta) = -1$. Thus, for integer $T > 1$,

$$S = \sum_{t=1}^{T} e^{2\pi ikt/T} = 0 \qquad k = \pm 1, \pm 2, ... \pm (T - 1)$$

as stated in Section 2.1.

   This algebra has an interesting geometric interpretation.   The statement that $z^T = 1$ means z is a Tth root of one. One itself is always a solution, but there are $T - 1$ additional solutions. It is easily verified that $e^{2\pi it/T}$, $t = 1, 2, ...T$ are T distinct solutions. If graphed in the complex plane, these T points are equidistantly located on the unit circle. Simple geometry indicates that they sum to zero. As long as k is an integer with $0 < |k| < T$, the same argument works for the equivalent set of solutions $e^{2\pi ikt/T}$, $t = 1, 2, ...T$.