# Time Series Analysis:

# Theory, Data Analysis and Computation

*by*

*Edward J. Wegman*
*Center for Computational Statistics*
*George Mason University*
*Fairfax, VA 22030*

# Chapter 1

# Introduction

Time series analysis deals with the *probabilistic* and *structural* inference about a sequence of data evolving through time. Normally, we shall write $X_1, \ldots, X_n$ as the time series observations. Frequently, this will be a sample from a continuous stochastic process $X_t(\omega)$ so that $X_j$ really means $X_{j\Delta t}$ where $\Delta$t is the sampling interval. Our basic model conceptually is a stochastic process $X_t(\omega)$ where t is the time index and $\omega \in \Omega$, where $(\Omega, \mathcal{A}, P)$ is a probability space and P is a probability measure. In simple terms, we are interested in discovering P based on a time series $X_1, \ldots, X_n$. Fortunately, or unfortunately, this is a ridiculously hard problem, in general. Hence, we typically make simplifying assumptions. For the purposes of our discussion, we will usually only consider finite dimensional distributions, i.e.

$$F_{x_{t_1}\ldots x_{t_k}}(u_1, \ldots, u_k) = P(X_{t_1} \leq u_1, \ldots, X_{t_k} \leq u_k), \forall k, \forall t_1, \ldots, t_k, \forall u_1, \ldots, u_k$$

which is in the simple one dimensional case,

$$F_{x_j}(u_j) = P(X_j \leq u_j).$$

The well-known Kolmogorov Extension Theorem basically says that knowledge of all finite-dimensional distributions is sufficient to reconstruct the probability measure P. This theorem is well beyond the scope of our course, however, and hence will not be discussed any further.

We actually go even further in making assumptions and assume that the finite dimensional distributions are all multivariate normal (or Gaussian). Since this is the case, it is sufficient to know only the first and second moments because the first and second moments of a Gaussian distribution is necessary and sufficient to characterize that distribution. We let $\mu_t = EX_t$ and $\gamma_{t_1,t_2} = \text{cov}(X_{t_1}, X_{t_2})$. If we know $\mu_t$ and $\gamma_{t_1,t_2}$, then we have enough information to reconstruct all of the finite dimensional distributions and hence the probability measure, P.

We actually begin our discussion with even one more simplification. We basically want our process to be invariant with a time shift, i.e. time homogeneous so that we will assume, at least initially, that the probability structure does not shift with time. We write this as follows:

(1.1) $$F_{x_{t_1}\ldots x_{t_k}}(t_1, \ldots, t_k) = F_{x_{t_1+\tau}\ldots x_{t_k+\tau}}(t_1, \ldots, t_k).$$

for every $t_1, \ldots, t_k$, for every $i_1, \ldots, i_k$, for every k and for every $\tau$. The parameter $\tau$ is the constant time shift. If this equation holds we say the process $X_t$ is *strictly stationary*.

A corollary to our Gaussian assumption is that strictly stationary implies means and covariances are invariant with time shifts. In particular,

(1.2) $$\mu_t = \mu_{t+\tau}$$

and

(1.3) $$\gamma_{t_1,t_2} = \gamma_{t_1+\tau,t_2+\tau}.$$

Thus if we let $\tau = -t$ in equation (1.2), we have

$$\mu_t = \mu_0, \text{ a constant,}$$

and if we let $\tau = t_2$ in equation (1.3), we have

$$\gamma_{t_1,t_2} = \gamma_{t_1-t_2,0} \overset{def}{=} \gamma_{t_1-t_2}.$$

Thus, the mean is constant; we will usually assume $\mu_0 = 0$ for convenience. Moreover, the covariance depends only on the difference between the indices, called the *lag*. $\gamma_{t_1-t_2}$ is called the *autocovariance* or *lag covariance*. A process which has a constant first moment and a covariance function which depends only on the lag is said to be *weakly stationary* or *wide-sense stationary*. A process may be strictly stationary and not weakly stationary if it has no second-order moments. Conversely, a process may be weakly stationary, but not strictly stationary is say its third moments change with time. Usually there is no confusion since under Gaussian distribution assumptions, these two notions are equivalent. In this case we simply say the process is *stationary*.

The problem of estimating P has been reduced to the much easier problem of estimating $\gamma_\tau$. This is the *probabilistic* or *statistical* inference problem. We are also interested in *structural* inference, i.e. finding the functional structure of the $X_t$ process. We can do this in the *frequency domain* or the *time domain*.

In the frequency domain, we consider the function,

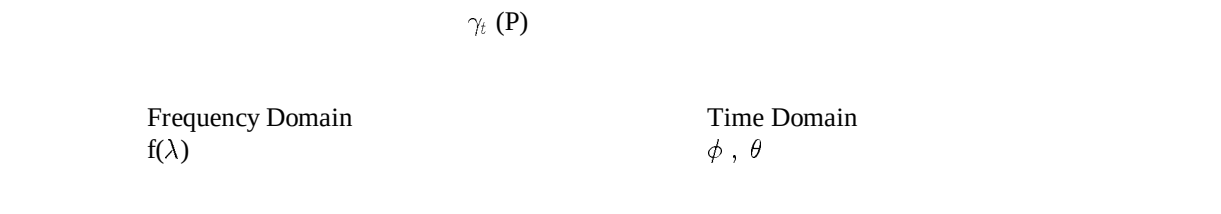
$$f(\lambda) = \int e^{-q^i\lambda\tau} \gamma_\tau \, d\tau \tag{1.4}$$

This is called the *spectral density* or *power spectral density*. The spectral density is the Fourier transform of the covariance function and hence carries the same information as the covariance function.

The time domain models, usually assumed to be linear and stationary, are of the form:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}, \tag{1.5}$$

where $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ are parameters and $\epsilon_t$ is a white noise. This is called the autoregressive-moving average model (ARMA) and is the most general *linear* model of a stationary stochastic process. As we shall later see, the parameters $\phi = (\phi_1, \ldots, \phi_p)$ and $\theta = (\theta_1, \ldots, \theta_q)$ determine and are determined by the covariance function. Moreover, we will see that we can write the spectral density in terms of $\phi$ and $\theta$ so that we have a fundamental triad given in the figure below.




Basically given one corner, we can derive the other two. Hence, under the stationary, Gaussian assumptions, analyses in the probability domain, the frequency domain, or the time domain are all basically equivalent and fundamentally dependent upon the mean and covariance structure of the process. The nature of the problem and what we are trying to infer will generally lead us to prefer one mode over another.

Generally speaking, engineering applications tend to succeed somewhat better with a frequency domain model while business and economic applications tend to be somewhat better represented by time domain models. This, of course, is *not* a hard and fast rule by any means. We will consider both frequency domain and time domain models.

# Chapter 2

# Frequency Models

The major focus of the first part of our discussion is structural inference for time series in the frequency domain. Historically, one of the earliest attempts to model this time series structure is known as the *method of hidden periodicities*. The premise for such a techniques is that the time series may be modeled by a linearly superimposed sum of sinusoids together with some Gaussian errors. We will develop such a model in several stages.

**2.1. A Single Sinusoid in Noise.** Consider the model, $X_t = R\cos(\lambda t + \phi)$, $\lambda$ real-valued is called the (angular) frequency, $\phi \in [-\pi, \pi]$ is called the phase, $R \geq 0$ is called the amplitude. The stochastic model is

$$X_t = \mu + R\cos(\lambda t + \phi) + \epsilon_t. \tag{2.1.1}$$

Recall trigonometric identity

$$\cos(\alpha+\beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta). \tag{2.1.2}$$

We may use (2.1.2) to obtain

$$R\cos(\lambda t + \phi) = R\cos(\lambda t)\cos(\phi) - R\sin(\lambda t)\sin(\phi). \tag{2.1.3}$$

Equivalently then, we may rewrite (2.1.1)

$$X_t = \mu + A\cos(\lambda t) + B\sin(\lambda t) + \epsilon_t \tag{2.1.4}$$

where,

$$A = R\cos(\phi),\ B = -R\sin(\phi).$$

**Least Squares Solution with 0 Mean.**

We use least squares for the estimation of A and B. Set $\mu = 0$ temporarily and form the sum of squares (SS).

$$SS = \sum_{t=0}^{n-1} (X_t - A\cos(\lambda t) - B\sin(\lambda t))^2 = \sum_{t=0}^{n-1} \epsilon_t^2. \tag{2.1.5}$$

Taking partial derivatives with respect to A and B, we obtain

$$\frac{\partial SS}{\partial A} = \sum_{t=0}^{n-1} 2\,(X_t - A\cos(\lambda t) - B\sin(\lambda t))\,(-\cos(\lambda t))$$

$$= -\sum_{t=0}^{n-1} 2\cos(\lambda t)\,(X_t - A\cos(\lambda t) - B\sin(\lambda t))$$

and

$$\frac{\partial SS}{\partial B} = \sum_{t=0}^{n-1} 2\,(X_t - A\cos(\lambda t) - B\sin(\lambda t))\,(-\sin(\lambda t))$$

$$= \sum_{t=0}^{n-1} -2\sin(\lambda t)\,(X_t - A\cos(\lambda t) - B\sin(\lambda t)).$$

Setting partials $= 0$ and collecting coefficients,

$$A\sum_{t=0}^{n-1}\cos^2(\lambda t) + B\sum_{t=0}^{n-1}\cos(\lambda t)\sin(\lambda t) - \sum_{t=0}^{n-1}\cos(\lambda t)\,X_t = 0$$

$$A\sum_{t=0}^{n-1}\cos(\lambda t)\sin(\lambda t) + B\sum_{t=0}^{n-1}\sin^2(\lambda t) - \sum_{t=0}^{n-1}\sin(\lambda t)\,X_t = 0$$

or rewriting,

$$A\,a + B\,b - \sum_{t=0}^{n-1} X_t\cos(\lambda t) = 0$$

$$A\,b + B\,c - \sum_{t=0}^{n-1} X_t\sin(\lambda t) = 0.$$

Here we are letting $a = \sum_{t=0}^{n-1}\cos^2(\lambda t)$, $b = \sum_{t=0}^{n-1}\cos(\lambda t)\sin(\lambda t)$ and $c = \sum_{t=0}^{n-1}\sin^2(\lambda t)$.

Solving for B, we have

$$A\,ab + B\,b^2 - b\sum_{t=0}^{n-1} X_t\cos(\lambda t) = 0$$

and

$$A\,ab + B\,ac - a\sum_{t=0}^{n-1} X_t\sin(\lambda t) = 0.$$

So that

(2.1.6)
$$\hat{B} = \frac{a\sum_{t=0}^{n-1} X_t\sin(\lambda t) - b\sum_{t=0}^{n-1} X_t\cos(\lambda t)}{ac - b^2}$$

Similarly,

$$A\,ac + B\,bc - c\sum_{t=0}^{n-1} X_t\cos(\lambda t) = 0$$

$$A\,b^2 + B\,bc \;-\; b\sum_{t=0}^{n-1} X_t \sin(\lambda t) \;=\; 0.$$

Subtracting top from bottom,

$$A\,(b^2 \;-\; ac) + c\sum_{t=0}^{n-1} X_t \cos(\lambda t) \;-\; b\sum_{t=0}^{n-1} X_t \sin(\lambda t) \;=\; 0$$

which yields

(2.1.7)
$$\hat{A} \;=\; \frac{c\sum_{t=0}^{n-1} X_t \cos(\lambda t) - b\sum_{t=0}^{n-1} X_t \sin(\lambda t)}{ac - b^2}.$$

Thus writing out fully,

$$\hat{A} \;=\; \frac{\sum_{t=0}^{n-1}\sin^2(\lambda t)\sum_{t=0}^{n-1} X_t \cos(\lambda t) - \sum_{t=0}^{n-1}\sin(\lambda t)\cos(\lambda t)\sum_{t=0}^{n-1} X_t \sin(\lambda t)}{D}$$

(2.1.8) and

$$\hat{B} \;=\; \frac{\sum_{t=0}^{n-1}\cos^2(\lambda t)\sum_{t=0}^{n-1} X_t \sin(\lambda t) - \sum_{t=0}^{n-1}\sin(\lambda t)\cos(\lambda t)\sum_{t=0}^{n-1} X_t \cos(\lambda t)}{D}$$

where

$$D = ac - b^2 = \sum_{t=0}^{n-1}\sin^2(\lambda t)\sum_{t=0}^{n-1}\cos^2(\lambda t) - \Big(\sum_{t=0}^{n-1}\sin(\lambda t)\cos(\lambda t)\Big)^2.$$

**Closed Form of the Least Squares Solution with 0 Mean**

To develop a closed form least squares solution, we may consider the identity,

$$\cos(\alpha)\cos(\beta) \;=\; \tfrac{1}{2}\{\cos(\alpha + \beta) + \cos(\alpha \;-\; \beta)\}.$$

Letting $\alpha = \beta = \lambda t$,

$$\sum_{t=0}^{n-1}\cos^2(\lambda t) \;=\; \sum_{t=0}^{n-1}\cos(\lambda t)\cos(\lambda t) \;=\; \tfrac{1}{2}\sum_{t=0}^{n-1}\{\cos(2\lambda t) + \cos(0)\}$$

$$= \tfrac{1}{2}\sum_{t=0}^{n-1}\cos(2\lambda t) + \tfrac{n}{2}$$

To find a closed form expression for $\sum_{t=0}^{n-1} \cos(2\lambda t)$, we recall the fundamental identity,

$$\cos(\alpha t) = \frac{e^{i\alpha t} + e^{-i\alpha t}}{2}$$

so that

$$\sum_{t=0}^{n-1} \cos(\alpha t) = \frac{1}{2}\left\{ \sum_{t=0}^{n-1} e^{i\alpha t} + \sum_{t=0}^{n-1} e^{-i\alpha t} \right\}.$$

Each of these is a geometric series, so that,

$$\sum_{t=0}^{n-1} e^{i\alpha t} = \frac{1 - e^{i\alpha n}}{1 - e^{i\alpha}} \quad \text{and} \quad \sum_{t=0}^{n-1} e^{-i\alpha t} = \frac{1 - e^{-i\alpha n}}{1 - e^{-i\alpha}}$$

Combining these last two expressions, we obtain

$$\sum_{t=0}^{n-1} \cos(\alpha t) = \frac{1}{2}\left\{ \frac{1 - e^{i\alpha n}}{1 - e^{i\alpha}} + \frac{1 - e^{-i\alpha n}}{1 - e^{-i\alpha}} \right\}$$

Factoring out $e^{\frac{i\alpha n}{2}}$ in numerator of first term, $e^{\frac{i\alpha}{2}}$ in denominator of the first, $e^{-\frac{i\alpha n}{2}}$ in the numerator of the second term and finally $e^{-\frac{i\alpha}{2}}$ in the denominator of the second, we have

$$\sum_{t=0}^{n-1} \cos(\alpha t) = \frac{1}{2}\left\{ \frac{e^{\frac{i\alpha n}{2}}\left\{ e^{-\frac{i\alpha n}{2}} - e^{\frac{i\alpha n}{2}} \right\}}{e^{\frac{i\alpha}{2}}\left\{ e^{-\frac{i\alpha}{2}} - e^{\frac{i\alpha}{2}} \right\}} + \frac{e^{-\frac{i\alpha n}{2}}\left\{ e^{\frac{i\alpha n}{2}} - e^{-\frac{i\alpha n}{2}} \right\}}{e^{-\frac{i\alpha}{2}}\left\{ e^{\frac{i\alpha}{2}} - e^{-\frac{i\alpha}{2}} \right\}} \right\}.$$

So that

$$\sum_{t=0}^{n-1} \cos(\alpha t) = \left\{ \frac{e^{\frac{i\alpha(n-1)}{2}} + e^{\frac{-i\alpha(n-1)}{2}}}{2} \right\} \frac{\left\{ e^{-\frac{i\alpha n}{2}} - e^{\frac{i\alpha n}{2}} \right\}}{\left\{ e^{-\frac{i\alpha}{2}} - e^{\frac{i\alpha}{2}} \right\}}$$

or

$$\sum_{t=0}^{n-1} \cos(\alpha t) = \cos\left(\alpha \frac{(n-1)}{2}\right) \frac{\sin\left(\frac{n\alpha}{2}\right)}{\sin\left(\frac{\alpha}{2}\right)}.$$

Thus letting $\alpha = 2\lambda$

$$\sum_{t=0}^{n-1} \cos(2\lambda t) = \cos\left((n-1)\lambda\right) \frac{\sin(n\lambda)}{\sin(\lambda)}.$$

Finally we have

(2.1.9)
$$\sum_{t=0}^{n-1} \cos^2(\lambda t) = \frac{1}{2} \cos\left((n-1)\lambda\right) \frac{\sin(n\lambda)}{\sin(\lambda)} + \frac{n}{2}$$

$$= \frac{n}{2}\left\{1 + \frac{\sin(n\lambda)}{n\sin(\lambda)} \cos\left((n-1)\lambda\right)\right\}$$

$$= \frac{n}{2}\{1 + D_n(\lambda) \cos\left((n-1)\lambda\right)\}.$$

The function, $D_n(\lambda) = \frac{\sin(n\lambda)}{n\sin(\lambda)}$, is called the Dirichlet Kernel. By similar arguments,

(2.1.10)
$$\sum_{t=0}^{n-1} \cos(\lambda t)\sin(\lambda t) = \frac{n}{2} D_n(\lambda) \sin\left((n-1)\lambda\right)$$

and

(2.1.11)
$$\sum_{t=0}^{n-1} \sin^2(\lambda t) = \frac{n}{2}\{1 - D_n(\lambda) \cos\left((n-1)\lambda\right)\}.$$

Notice as verification

$$\sum_{t=0}^{n-1} \sin^2(\lambda t) + \sum_{t=0}^{n-1} \cos^2(\lambda t) = \frac{n}{2}\{1 + D_n(\lambda) \cos\left((n-1)\lambda\right)\}$$

$$+ \frac{n}{2}\{1 - D_n(\lambda)\cos\left((n-1)\lambda\right) = \frac{n}{2} + \frac{n}{2} = n.$$

Now we can write $\hat{A}$ and $\hat{B}$ in terms of the Dirichlet Kernel in closed form by substituting (2.1.9) through (2.1.11) in (2.1.8). Now we have defined a, b and c as follows.

$$a = \frac{n}{2}\left\{1 + D_n(\lambda)\cos\left((n-1)\lambda\right)\right\},$$

(2.1.12)
$$b = \frac{n}{2} D_n(\lambda)\sin\left((n-1)\lambda\right),$$

and

$$c = \frac{n}{2}\left\{1 - D_n(\lambda)\cos\left((n-1)\lambda\right)\right\}.$$

**Least Squares Solutions in Amplitude-Phase Form with 0 Mean**

Next recall that $A = R\cos(\phi)$ , $B = -R\sin(\phi)$ so that

$$A^2 + B^2 = R^2(\cos^2(\phi) + \sin^2(\phi)) = R^2.$$

Therefore $\hat{R} = \sqrt{\hat{A}^2 + \hat{B}^2}$. Notice also that $-\frac{B}{A} = \frac{R\sin(\phi)}{R\cos(\phi)} = \tan(\phi)$.

We must be careful in calculating $\hat{\phi}$ since the arctangent is not unique. There are several cases to be considered.

Case 1. If A is positive ($A > 0$), then $\cos(\phi) > 0$ which implies $-\frac{\pi}{2} < \phi < \frac{\pi}{2}$. Thus we are on a principal axis, so that,

$$\phi = \arctan(-\tfrac{B}{A}).$$

Case 2. If $A < 0$, $B > 0$, then $\cos(\phi) < 0$ so that $\phi$ is between $-\frac{\pi}{2}$ and $-\frac{3\pi}{2}$. Also $B > 0$ implies that $\sin(\phi) < 0$ which in turn implies $\phi \in (-\pi, -\frac{\pi}{2})$. In this case

$$\phi = \arctan(-\tfrac{B}{A}) - \pi.$$

Case 3. If $A < 0$, $B < 0$ implies $\cos(\phi) < 0$ and $B < 0$ implies $\sin(\phi) > 0$. In this case, $\phi \in (\frac{\pi}{2}, \pi)$ so that

$$\phi = \arctan(-\tfrac{B}{A}) + \pi.$$

Case 4. If $B = 0$, then $\phi = 0$.

Case 5. If $A = 0$, $B > 0$, then $\phi = -\frac{\pi}{2}$.

Case 6. If $A = 0$, $B < 0$, then $\phi = \frac{\pi}{2}$.

Thus we can determine $\hat{R}, \hat{\phi}$ from $\hat{A}$ and $\hat{B}$ which can in turn be estimated from the data.

**Least Squares Solution with Non-Zero Mean**

If $\mu$, the mean, is not equal to 0, when we may construct normal equations as follows. Consider again the sum of squares.

$$SS = \sum_{t=0}^{n-1} \epsilon_t^2 = \sum_{t=0}^{n-1} (X_t - \mu - A\cos(\lambda t) - B\sin(\lambda t))^2.$$

Taking partial derivatives and equating to 0 as before, we have

(2.1.13) $$\frac{\partial SS}{\partial A} = -2\sum_{t=0}^{n-1} \{((X_t - \mu) - A\cos(\lambda t) - B\sin(\lambda t))\cos(\lambda t)\} = 0$$

(2.1.14) $$\frac{\partial SS}{\partial B} = -2\sum_{t=0}^{n-1}\{((X_t-\mu) - A\cos(\lambda t) - B\sin(\lambda t))\sin(\lambda t)\} = 0$$

and

(2.1.15) $$\frac{\partial SS}{\partial \mu} = -2\sum_{t=0}^{n-1}\{(X_t-\mu) - A\cos(\lambda t) - B\sin(\lambda t)\} = 0.$$

Equations (2.1.13) and (2.1.14) are the same as our previous normal equations with $X_t - \mu$ substituted for $X_t$. Thus the solutions for $\hat{A}$ and $\hat{B}$ are

$$\hat{A} = \frac{\sum_{t=0}^{n-1}\sin^2(\lambda t)\sum_{t=0}^{n-1}(X_t-\hat{\mu})\cos(\lambda t) - \sum_{t=0}^{n-1}\sin(\lambda t)\cos(\lambda t)\sum_{t=0}^{n-1}(X_t-\hat{\mu})\sin(\lambda t)}{D}$$

(2.1.16) and

$$\hat{B} = \frac{\sum_{t=0}^{n-1}\cos^2(\lambda t)\sum_{t=0}^{n-1}(X_t-\hat{\mu})\sin(\lambda t) - \sum_{t=0}^{n-1}\sin(\lambda t)\cos(\lambda t)\sum_{t=0}^{n-1}(X_t-\hat{\mu})\cos(\lambda t)}{D}$$

where again

$$D = \sum_{t=0}^{n-1}\sin^2(\lambda t)\sum_{t=0}^{n-1}\cos^2(\lambda t) - (\sum_{t=0}^{n-1}\sin(\lambda t)\cos(\lambda t))^2.$$

Returning to (2.1.15), we have

$$\sum_{t=0}^{n-1} X_t - n\mu - \sum_{t=0}^{n-1} A\cos(\lambda t) - \sum_{t=0}^{n-1} B\sin(\lambda t) = 0.$$

Which we can solve for $\mu$ by

(2.1.17) $$\hat{\mu} = \bar{X} - \frac{\sum_{t=0}^{n-1}\left\{\hat{A}\cos(\lambda t) + \hat{B}\sin(\lambda t)\right\}}{n}.$$

Notice two dilemmas: 1. As formulated solution for $\hat{A}$ and $\hat{B}$ depends on $\hat{\mu}$ and similarly the solution for $\hat{\mu}$ depends an $\hat{A}$ and $\hat{B}$, and 2. Notice also that the estimator for $\mu$ is not $\bar{X}$. Thus as things stand now we have a recursive, but not direct solution for the case that $\mu \neq 0$. Remember that our first inclination would be to centralize the data by subtract $\bar{X}$, but (2.1.17) tells us that we may not do this either and have a least squares solution.

**The Matrix Formulation of the General Least Squares Problem**

Letting a, b and c have their previous meaning and now defining $d = \sum_{t=0}^{n-1}\cos(\lambda t)$ and $e = \sum_{t=0}^{n-1}\sin(\lambda t)$, we can write

$$A\,a + B\,b + \mu\,d \;-\; \sum_{t=0}^{n-1} Xt\cos(\lambda t) \;=\; 0$$

(2.1.18)
$$A\,b + B\,c + \mu\,e \;-\; \sum_{t=0}^{n-1} Xt\sin(\lambda t) \;=\; 0$$

and

$$A\,d + B\,e + \mu\,n \;-\; \sum_{t=0}^{n-1} Xt \;=\; 0.$$

In matrix form, we may write

(2.1.19)
$$\begin{bmatrix} a & b & d \\ b & c & e \\ d & e & n \end{bmatrix} \begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{\mu} \end{bmatrix} = \begin{bmatrix} \sum_{t=0}^{n-1} Xt\cos(\lambda t) \\ \sum_{t=0}^{n-1} Xt\sin(\lambda t) \\ \sum_{t=0}^{n-1} Xt \end{bmatrix}.$$

Taking inverse matrix and solving for the unknown parameters, we have

(2.1.20)
$$\begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{\mu} \end{bmatrix} = \begin{bmatrix} a & b & d \\ b & c & e \\ d & e & n \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=0}^{n-1} Xt\cos(\lambda t) \\ \sum_{t=0}^{n-1} Xt\sin(\lambda t) \\ \sum_{t=0}^{n-1} Xt \end{bmatrix}.$$

This is a linear model.

**A Computationally Quick and Dirty Approximation**

While the matrix formulation gives an exact solution, it does involve a matrix inversion. This is really not a serious impediment to calculation of a solution.
Nonetheless, a simple quick approximation may be given. Consider the Dirichlet Kernel, $D_n(\lambda) = \frac{\sin(n\lambda)}{n\sin(\lambda)}$. Now

(2.1.21) $$| \; n\, D_n(\lambda) \; | \;=\; \left| \frac{\sin(n\lambda)}{\sin(\lambda)} \right| \;\leq\; \left| \frac{1}{\sin(\lambda)} \right|$$

for $\lambda$ bounded away from 0 and $\pi$.

Thus

$$\sum_{t=0}^{n-1} \cos^2(\lambda t) \doteq n/2,$$

(2.1.22) $$\sum_{t=0}^{n-1} \sin(\lambda t)\cos(\lambda t) \doteq 0,$$

and

$$\sum_{t=0}^{n-1} \sin^2(\lambda t) \doteq n/2.$$

Thus we obtain a quick approximate solution as follows

$$\hat{A} = \frac{2}{n} \sum_{t=0}^{n-1} (X_t - \hat{\mu}) \cos(\lambda t),$$

(2.1.23) $$\hat{B} = \frac{2}{n} \sum_{t=0}^{n-1} (X_t - \hat{\mu}) \sin(\lambda t)$$

and

$$\hat{\mu} = \bar{X}.$$

**Least Squares Estimation of the Optimal Frequency, $\lambda$, in Model (2.1.1)**

Using this formulation of the quick and dirty estimates of $\hat{A}$, $\hat{B}$ and $\hat{\mu}$, we may perform a calculation for the residual sum of squares,

$$\text{SS (residuals)} = \sum_{t=0}^{n-1} (X_t - \hat{\mu} - \hat{A} \cos \lambda t - \hat{B} \sin \lambda t)^2$$

$$= \sum_{t=0}^{n-1} (X_t - \hat{\mu})^2 - 2 \sum_{t=0}^{n-1} (X_t - \hat{\mu}) (\hat{A} \cos(\lambda t) + \hat{B} \sin(\lambda t)) + \sum_{t=0}^{n-1} (\hat{A} \cos(\lambda t) + B \sin(\lambda t))^2$$

$$= \sum_{t=0}^{n-1} (X_t - \hat{\mu})^2 - 2n\hat{A}^2 - 2n\hat{B}^2 + \hat{A}^2 \sum_{t=0}^{n-1} \cos^2(\lambda t) + \hat{B}^2 \sum_{t=0}^{n-1} \sin^2(\lambda t)$$

$$+ 2\hat{A}\hat{B} \sum_{t=0}^{n-1} \sin(\lambda t)\cos(\lambda t)$$

$$\doteq \sum_{t=0}^{n-1} (X_t - \hat{\mu})^2 - n\hat{A}^2 - n\hat{B}^2 + \hat{A}^2\left(\tfrac{n}{2}\right) + \hat{B}^2\left(\tfrac{n}{2}\right) + 0.$$

Therefore,

$$\text{SS (residuals)} \doteq \sum_{t=0}^{n-1} (X_t - \hat{\mu})^2 - \tfrac{n}{2}(\hat{A}^2 + \hat{B}^2).$$

We may restructure this equation slightly into

$$\text{SS(corrected)} = \sum_{t=0}^{n-1} (X_t - \hat{\mu})^2 \doteq \text{SS (residuals) + SS (due to model at } \lambda)$$

where we define

$$\text{SS (due to model at } \lambda) = \tfrac{n}{2}(\hat{A}^2 + \hat{B}^2) = \tfrac{n}{2}\hat{R}^2.$$

We recall that $R^2$ is amplitude squared of the cosine model, (2.1.1) and note that we can regard $R^2$ as a function of the frequency, $R^2(\lambda)$. This suggests that we can also estimate $\lambda$. Thus, the least squares estimate of $\lambda$ is $\hat{\lambda}$ that minimizes SS(residuals) or equivalently maximizes $R^2(\lambda)$. $R^2(\lambda)$, it turns out, plays an interesting role in frequency domain analysis. As we shall see later, $I_n(\lambda) = \frac{n}{8\pi}\hat{R}^2(\lambda)$ is called the *periodogram.*

### 2.2 The Generalized Hidden Periodicities Model

We now consider the generalized model which is the sum of sinusoidal components.

(2.2.1) $$X_t = \mu + \sum_{j=1}^{m} \{A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t)\} + \epsilon_t \text{ , } t = 0, \ldots, n-1$$

In this case,

(2.2.2) $$SS = \sum_{t=0}^{n-1} \epsilon_t^2 = \sum_{t=0}^{n-1} \Big(X_t - \mu - \sum_{j=1}^{m} \{A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t)\}\Big)^2.$$

Taking derivatives and equating to 0, we find we have $2m + 1$ simultaneous equations

$$\frac{\partial SS}{\partial A_k} = 2\sum_{t=0}^{n-1} \Big(X_t - \mu - \sum_{j=1}^{m} \{A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t)\}\Big)\cos(\lambda_k t) = 0, \; k = 1, \ldots, m$$

$$\frac{\partial SS}{\partial B_k} = 2\sum_{t=0}^{n-1} \Big(X_t - \mu - \sum_{j=1}^{m} \{A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t)\}\Big)\sin(\lambda_k t) = 0, \; k = 1, \ldots, m$$

and

$$\frac{\partial SS}{\partial \mu} = 2\sum_{t=0}^{n-1} \Big(X_t - \mu - \sum_{j=1}^{m} \{A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t)\}\Big) = 0.$$

Letting $a_{jk}$, $b_{jk}$, $c_{jk}$, $d_j$ and $e_j$ be defined as follows:

$$a_{jk} = \sum_{t=0}^{n-1} \cos(\lambda_j t) \cos(\lambda_k t)$$

$$b_{jk} = \sum_{t=0}^{n-1} \sin(\lambda_j t) \cos(\lambda_k t)$$

(2.2.3)
$$c_{jk} = \sum_{t=0}^{n-1} \sin(\lambda_j t) \sin(\lambda_k t)$$

$$d_j = \sum_{t=0}^{n-1} \cos(\lambda_j t)$$

and

$$e_j = \sum_{t=0}^{n-1} \sin(\lambda_j t).$$

We may then rewrite the normal equations as

$$\sum_{j=1}^{m} A_j a_{jk} + \sum_{j=1}^{m} B_j b_{jk} + \mu d_k - \sum_{t=0}^{n-1} X_t \cos(\lambda_k t) = 0, \quad k = 1, \ldots, m$$

(2.2.4)
$$\sum_{j=1}^{m} A_j b_{kj} + \sum_{j=1}^{m} B_j c_{jk} + \mu e_k - \sum_{t=0}^{n-1} X_t \sin(\lambda_k t) = 0, \quad k = 1, \ldots, m$$

and

$$\sum_{j=1}^{m} A_j d_j + \sum_{j=1}^{m} B_j e_j + \mu n - \sum_{t=0}^{n-1} X_t = 0.$$

In matrix form we can write

$$\underset{\sim}{a} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & & & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{mm} \end{bmatrix}$$

$$\underset{\sim}{b} = \begin{bmatrix} b_{11} & b_{21} & \cdots & b_{m1} \\ b_{12} & b_{22} & \cdots & b_{m2} \\ \vdots & & & \vdots \\ b_{1m} & b_{2m} & \cdots & b_{mm} \end{bmatrix}$$

$$\underset{\sim}{c} = \begin{bmatrix} c_{11} & c_{21} & \cdots & c_{m1} \\ c_{12} & c_{22} & \cdots & c_{m2} \\ \vdots & & & \vdots \\ c_{1m} & c_{2m} & \cdots & c_{mm} \end{bmatrix}$$

(2.2.5)

$$\underset{\sim}{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{bmatrix}, \qquad \underset{\sim}{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}$$

$$\underset{\sim}{A} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix}, \qquad \underset{\sim}{B} = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_m \end{bmatrix}$$

and finally,

$$\underset{\sim}{S} = \begin{bmatrix} \sum_{t=0}^{n-1} Xt \cos(\lambda_1 t) \\ \vdots \\ \sum_{t=0}^{n-1} Xt \cos(\lambda_m t) \\ \sum_{t=0}^{n-1} Xt \sin(\lambda_1 t) \\ \vdots \\ \sum_{t=0}^{n-1} Xt \sin(\lambda_m t) \\ \sum_{t=0}^{n-1} Xt \end{bmatrix}. \tag{2.2.6}$$

In block matrix form, we may then write

$$\begin{bmatrix} \underset{\sim}{a} & \underset{\sim}{b} & \underset{\sim}{d} \\ \underset{\sim}{b}^T & \underset{\sim}{c} & \underset{\sim}{e} \\ \underset{\sim}{d}^T & \underset{\sim}{e}^T & n \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{A}} \\ \hat{\underset{\sim}{B}} \\ \hat{\mu} \end{bmatrix} = \underset{\sim}{S} \tag{2.2.7}$$

Thus taking inverses, we may write

$$\begin{bmatrix} \hat{\underset{\sim}{A}} \\ \hat{\underset{\sim}{B}} \\ \hat{\mu} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{a} & \underset{\sim}{b} & \underset{\sim}{d} \\ \underset{\sim}{b}^T & \underset{\sim}{c} & \underset{\sim}{e} \\ \underset{\sim}{d}^T & \underset{\sim}{e}^T & n \end{bmatrix}^{-1} \underset{\sim}{S}. \tag{2.2.8}$$

**Closed form expressions for $a_{ij}$, $b_{ij}$ and $c_{ij}$.**

It is clear that expressions of the form $a_{jk} = \sum_{j=0}^{n-1} \cos(\lambda j t) \cos(\lambda_k t)$ will play a critical role in evaluating equations (2.2.8). Let us temporarily divert our attentions to evaluating expressions of this sort. Let us restrict our attention to $0 \leq \lambda_j \leq 2\pi$ and we will eventually want to consider $\lambda_j = 2\pi$ (j/n), j = 0, 1, … , n – 1. Thus just as we consider sampled times j = 0 , … , n – 1, we consider sampled frequencies $\lambda_j$, j = 0 , … , n – 1.

Recall the identity we used earlier

$$\cos(a)\cos(b) = \tfrac{1}{2}\{\cos(a+b) + \cos(a-b)\}.$$

We may thus write

$$\sum_{t=0}^{n-1}\cos(\lambda_j t)\cos(\lambda_k t) \;=\; \tfrac{1}{2}\Big\{\sum_{t=0}^{n-1}\cos\Big((\lambda_j+\lambda_k)\,t\Big) \;+\; \sum_{t=0}^{n-1}\cos\Big((\lambda_j-\lambda_k)\,t\Big).$$

Using the earlier closed form solution and the Dirichlet formula, we may write

$$\sum_{t=0}^{n-1}\cos(\lambda_j t)\cos(\lambda_k t) \;=\; \tfrac{1}{2}\cos\Big(\tfrac{(n-1)(\lambda_j+\lambda_k)}{2}\Big)\; n\; D_n\Big(\tfrac{\lambda_j+\lambda_k}{2}\Big)$$

$$+\;\tfrac{1}{2}\cos\Big(\tfrac{(n-1)(\lambda_j-\lambda_k)}{2}\Big)\; n\; D_n\Big(\tfrac{\lambda_j-\lambda_k}{2}\Big)$$

where $D_n(x) \;=\; \frac{\sin(nx)}{n\sin(x)}$.

Other cross-product terms may be derived in a similar manner. If, in particular, $\lambda_j = 2\pi j/n$, then $\lambda_j + \lambda_k = 2\pi(j+k)/n$. If $j \neq k \bmod (n)$, then the first Dirichlet function $D_n(\pi(j+k)/n)$ has a numerator equal to $\sin\Big(\pi(j+k)\Big) \;=\; 0$ since $j+k$ is always an integer. In the same way, $D_n(\pi(j-k)/n)$ has a numerator equal to $\sin\Big(\pi(j-k)\Big) \;=\; 0$ since $j-k$ is also always an integer. If $j = k \bmod(n)$, then $k = j + nm$ for some integer m. Thus $\cos(\lambda_k t) \;=\; \cos(2\pi kt/n) \;=\; \cos\Big(2\pi(j+nm)t/n\Big) \;=\; \cos(2\pi jt/n) \;=\; \cos(\lambda_j t)$. Since we have $\lambda_j \;=\; 2\pi j/n$

(2.2.9) $$\sum_{t=0}^{n-1}\cos(\lambda_j t)\cos(\lambda_k t) \;=\; \sum_{t=0}^{n-1}\cos^2(\lambda_j t) \;=\; \tfrac{n}{2}\{1 + D_n(2\pi j/n)\cos\Big((n-1)\lambda_j\Big).$$

Again $D_n(2\pi j/n)$ has a numerator equal to $\sin(2\pi j) \;=\; 0$. Hence for $\lambda_j = 2\pi j/n$

(2.2.10) $$\sum_{t=0}^{n-1}\cos(\lambda_j t)\cos(\lambda_k t) \;=\; \tfrac{n}{2}.$$

Finally we may conclude that for $\lambda_j = 2\pi j/n$,

(2.2.11) $$a_{jk} = \begin{cases} n/2, & j=k \\ 0, & j\neq k. \end{cases}$$

Similarly,

(2.2.12) $$b_{jk} \;=\; 0, \forall j, k$$

and

(2.2.13) $$c_{jk} = \begin{cases} n/2, & j=k \\ 0, & j\neq k. \end{cases}$$

Finally we note that for $\lambda_j = 2\pi j/n$, $d_j = e_j = 0$.

**Generalized Quick and Dirty Methodology**

Based on substituting the results listed above as (2.2.11) through (2.2.13) into equation (2.2.8), it is not difficult to see that the quick and dirty formulae we had for the single sinusoid case generalizes directly to the multiple sinusoid case. That in particular the matrix in equation (2.2.8) becomes a diagonal matrix and the set of equations decouples into

$$\hat{A}_j \;=\; \tfrac{2}{n}\sum_{t=0}^{n-1}(X_t - \hat{\mu})\;\cos(\lambda_j t), \quad 0 < j \;\leq\; n-1$$

and

$$\hat{B}_j = \frac{2}{n} \sum_{t=0}^{n-1} (X_t - \hat{\mu}) \sin(\lambda_j t), \quad 0 < j \leq n-1$$

and because of symmetry

$$\hat{\mu} = \hat{A}_0 = \frac{1}{n} \sum_{t=0}^{n-1} X_t = \bar{X}$$

which are as before the quick and dirty estimators.

# Chapter 3
# Fourier Transforms and Harmonic Analysis

With the basic theory of frequency domain models now in hand, it is important to develop diagnostic tools to structure these models. One of the most powerful of these diagnostic tools is harmonic analysis which may be intuitively thought of as the decomposition of a time series into its sinusoidal components. Roughly speaking, our general frequency domain model (2.2.1) suggests that we conceive of a time series as the superposition of sinusoidal components. The determination of which of these frequency components, $\lambda_j$, are the critical ones is the point harmonic analysis and the basic mathematical tool for doing this is the Fourier transform. We begin by considering complex sinusoids. Again we will be assuming throughout the next several sections that $\lambda_j = 2\pi j/n$ as we did in the last section.

**Complex Sinusoids**

Recall that $e^{i\alpha} = \cos(\alpha) + i \sin(\alpha)$ which is the so-called *complex sinusoid.* From this it is easy to see that $e^{-i\alpha} = \cos(-\alpha) + i \sin(-\alpha) = \cos(\alpha) - i \sin(\alpha)$ for $\alpha \in (0, \pi)$. From this it is easy to derive the fundamental identity,

(3.1.1) $$\cos(\alpha) = \frac{e^{i\alpha} + e^{-i\alpha}}{2}$$

by addition. Similarly, by subtraction,

(3.1.2) $$\sin(\alpha) = \frac{e^{i\alpha} - e^{-i\alpha}}{2i}$$

an identity we used earlier. These are called the Eulerian identities.

Now

$$\sum_{t=0}^{n-1} e^{i\lambda_j t} e^{-i\lambda_k t} = \sum_{t=0}^{n-1} e^{i(\lambda_j - \lambda_k)t}$$

$$= \sum_{t=0}^{n-1} \cos\left((\lambda_j - \lambda_k)t\right) + i \sum_{t=0}^{n-1} \sin\left((\lambda_j - \lambda_k)t\right).$$

By our earlier derivation of sums of series and cosines, we have

$$\sum_{t=0}^{n-1} e^{i(\lambda_j - \lambda_k)t} = \left\{ \cos\left(\frac{(n-1)(\lambda_j - \lambda_k)}{2}\right) + i \sin\left(\frac{(n-1)(\lambda_j - \lambda_k)}{2}\right) \right\} n D_n\left(\frac{(\lambda_j - \lambda_k)}{2}\right)$$

But $\lambda_j - \lambda_k = \frac{2\pi(j-k)}{n}$ so that, $\sin\left(\frac{n(\lambda_j - \lambda_k)}{2}\right) = \sin\left(\pi(j-k)\right) = 0$ since the sine of an integer multiple of $\pi$ is 0. Thus the ratio

$$n\, D_n\left(\frac{(\lambda_j-\lambda_k)}{2}\right) = \sin\left(\frac{n(\lambda_j-\lambda_k)}{2}\right) / \sin\left(\frac{\lambda_j-\lambda_k}{2}\right)$$

is 0 unless $\lambda_j - \lambda_k = 0$. In this case the ratio is undefined. Since this ratio is 0, it follows that,

$$\sum_{t=0}^{n-1} e^{i(\lambda_j-\lambda_k)t} = 0 \text{ unless } \lambda_j - \lambda_k = 0.$$

Now if $\lambda_j - \lambda_k = 0$, then

$$\sum_{t=0}^{n-1} e^{i(\lambda_j-\lambda_k)t} = \sum_{t=0}^{n-1} e^0 = \sum_{t=0}^{n-1} 1 = n.$$

Thus we have

(3.1.3)
$$\sum_{t=0}^{n-1} e^{i(\lambda_j-\lambda_k)t} = \begin{cases} n, & j=k \bmod(n) \\ 0, & \text{otherwise.} \end{cases}$$

This is called the *orthogonality property*.

**The Discrete Fourier Transform**

We now consider a time series, $X_0, \ldots, X_{n-1}$ and define

(3.1.4)
$$J_j = \frac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i\lambda_j t},\ j = 1, \ldots, n-1.$$

The set of possibly complex numbers $J_0,\ldots,J_{n-1}$ is the *discrete Fourier transform (DFT)* of $X_0,\ldots,X_{n-1}$. To give a bit of intuition to our interpretation of the DFT, suppose $X_t$ and $Y_t$ are two times series each with 0 mean. Then the expression,

$$\text{cov}(X, Y) = \frac{1}{n}\sum_{t=0}^{n-1} X_t\, Y_t$$

is the sample covariance of the two time series. Now if we take $Y_t = e^{-i\lambda_j t} = \cos(\lambda_j t) - i\sin(\lambda_j t)$ as the fundamental complex sinusoid of frequency $\lambda_j$, then $J_j$ measures the degree of "correlation" of $X_t$ with this fundamental sinusoid of frequency $\lambda_j$. Of course, this is not a true correlation in the straight statistical sense since $J_j$ may be complex valued. However, if we look at the magnitude of $J_j$, i.e. $|J_j|$ we do indeed get something very close to a correlation. We shall examine this in more detail later. Let us notice now, however, that $\lambda_{n-j} = 2\pi(n-j)/n = 2\pi - 2\pi j/n$. Thus $\exp(-i\lambda_{n-j}t) = \exp(-2\pi it)\exp(2\pi jt/n)$. Because $-2\pi t$ is an integer multiple of $2\pi$, we have $\exp(-2\pi it) = 1$ and $\exp(-i\lambda_{n-j}t) = \exp(i\lambda_j t)$. Thus

(3.1.5)
$$J_{n-j} = \frac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i\lambda_{n-j}t} = \frac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i(-\lambda_j)t} = \frac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i\lambda_{-j}t} = J_{-j}.$$

Moreover, if $X_t$ is real, then $X_t^* = X_t$ ($^*$ will be used to indicate complex conjugate). Thus, if $z = a + ib$, then $z^* = a - ib$. Consider then for real $X_t$

(3.1.6)
$$J_{n-j}^* = \left\{\frac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i\lambda_{-j}t}\right\}^* = \frac{1}{n}\sum_{t=0}^{n-1} X_t^*\, e^{i\lambda_{-j}t} = \frac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i\lambda_j t} = J_j$$

Combining (3.1.5) and (3.1.6), we have $J_{-j} = J_j^*$. This is called the *Hermetian or symmetry property*. Now let $X_t = Y_t + Z_t$. Then

$$J_{x,j} = \frac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i\lambda_j t} = \frac{1}{n}\sum_{t=0}^{n-1} (Y_t + Z_t)\, e^{-i\lambda_j t}$$

$$= \frac{1}{n}\sum_{t=0}^{n-1} Y_t\, e^{-i\lambda_j t} + \frac{1}{n}\sum_{t=0}^{n-1} Z_t\, e^{-i\lambda_j t} = J_{y,j} + J_{z,j}.$$

This is called the *linear or superposition* property. Next we claim that

(3.1.7) $$X_t = \sum_{j=0}^{n-1} J_j\, e^{i\lambda_j t} \text{ if } J_k = \frac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i\lambda_k t}.$$

We consider then

(3.1.8) $$\frac{1}{n}\sum_{t=0}^{n-1}\sum_{j=0}^{n-1} J_j\, e^{i\lambda_j t} e^{-i\lambda_k t} = \frac{1}{n}\sum_{j=0}^{n-1} J_j \sum_{t=0}^{n-1} e^{i(\lambda_j - \lambda_k)t}.$$

Now by (3.1.3)

(3.1.9) $$\frac{1}{n}\sum_{t=0}^{n-1}\sum_{j=0}^{n-1} J_j\, e^{i\lambda_j t} e^{-i\lambda_k t} = \frac{1}{n}\sum_{j=0}^{n-1} J_j \times \begin{cases} n, & j = k \bmod(n) \\ 0, & j \neq k \bmod(n) \end{cases} = J_k.$$

Now because $J_k = \frac{1}{n}\sum_{t=0}^{n-1} X_t e^{-i\lambda_k t}$, we may identify $X_t$ in (3.1.8) with $\sum_{j=0}^{n-1} J_j\, e^{i\lambda_j t}$. These are *Fourier transform pairs.*

An interesting connection may be made with model (2.2.1). Choose this model with $\mu = 0$ and recall the quick and dirty solution for the $A_j$ and the $B_j$. In particular

(3.1.10) $$A_j = \frac{2}{n}\sum_{t=0}^{n-1} X_t \cos(\lambda_j t), \quad 0 < j \leq n-1$$

and

(3.1.11) $$B_j = \frac{2}{n}\sum_{t=0}^{n-1} X_t \sin(\lambda_j t), \quad 0 < j \leq n-1.$$

Consider the discrete Fourier transform

$$J_j = \frac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i\lambda_j t} = \frac{1}{n}\sum_{t=0}^{n-1} X_t \cos(\lambda_j t) - \frac{i}{n}\sum_{t=0}^{n-1} X_t \sin(\lambda_j t).$$

Thus we have

(3.1.12) $$J_j = \frac{1}{2} A_j - i\,\frac{1}{2} B_j.$$

Consider

(3.1.13) $$|J_j|^2 = J_j J_j^* = \frac{1}{4}(A_j - i B_j)(A_j + i B_j) = \frac{1}{4}(A_j^2 - i^2 B_j^2)$$

$$= \tfrac{1}{4}(A_j^2 + B_j^2) = \tfrac{1}{4}R_j^2.$$

Thus the square modulus of the Fourier transform is related to the fundamental model. Recall from before that we had defined $R_j$ by the expression

(3.1.14) $$R_j \cos(\lambda t + \phi_j) = A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t).$$

Thus the quick and dirty computation of the solution to (2.2.1) can be computed in terms of the real and imaginary components of the Fourier transform of the time series, i.e. $A_j = 2$ real $(J_j)$ and $B_j = -2$ imaginary $(J_j)$.

The use of $J_j$ as a diagnostic tool is to be explored a bit further. Basically we gave a "correlation" type interpretation to $J_j$. Because $J_j \overset{*}{J_j} = R_j^2$ is the amplitude squared of the frequency component at $\lambda_j$. By plotting $R_j^2$, we can get an idea of power of frequency component at $\lambda_j$.

**Fourier Transforms for a Continuous Range of Frequencies**

A natural question concerns the extension of the concept of Fourier transforms to a continuous range of frequencies. Consider

(3.1.15) $$J(\lambda) = \frac{1}{n} \sum_{t=0}^{n-1} X_t \, e^{-i\lambda t}$$

as the natural extension of $J_k$. Let us consider the following facts. If $n = 0$,

$$\int_{-\pi}^{\pi} e^{i\lambda n} \, d\lambda = \int_{-\pi}^{\pi} d\lambda = 2\pi.$$

However, if $n \neq 0$,

$$\int_{-\pi}^{\pi} \cos(\lambda n) \, d\lambda = \frac{\sin(n\pi)}{n} - \frac{\sin(-n\pi)}{n} = \frac{0-0}{n} = 0.$$

Also $$\int_{-\pi}^{\pi} \sin(\lambda n) \, d\lambda = -\frac{\cos(n\pi) - \cos(-n\pi)}{n} = -\frac{1-1}{n} = 0.$$

Thus $$\int_{-\pi}^{\pi} e^{i\lambda n} \, d\lambda = 0.$$

Combining these results we have

(3.1.16) $$\int_{-\pi}^{\pi} e^{i\lambda n} \, d\lambda = \begin{cases} 2\pi, & n=0 \\ 0, & n \neq 0 \text{ and an integer} \end{cases}.$$

This result is analogous to the orthogonality result in the discrete case. We may derive

(3.1.17) $$X_t = \frac{n}{2\pi} \int_{-\pi}^{\pi} J(\lambda) \, e^{i\lambda t} \, d\lambda, \quad 0 \leq t < n$$

so that $J(\lambda)$ and $X_t$ are Fourier transform pairs as before. To see this consider,

$$\tfrac{n}{2\pi}\int_{-\pi}^{\pi} J(\lambda)\, e^{i\lambda t}\, d\lambda \;=\; \tfrac{n}{2\pi}\int_{-\pi}^{\pi} \tfrac{1}{n}\sum_{u=0}^{n-1} X_u\, e^{-i\lambda u}\, e^{i\lambda t}\, d\lambda$$

by definition. Re-ordering,

$$\tfrac{n}{2\pi}\int_{-\pi}^{\pi} J(\lambda)\, e^{i\lambda t}\, d\lambda = \tfrac{1}{2\pi}\sum_{u=0}^{n-1} X_u \int_{-\pi}^{\pi} e^{-i\lambda(t-u)}\, d\lambda.$$

But by (3.1.16),

$$\tfrac{n}{2\pi}\int_{-\pi}^{\pi} J(\lambda)\, e^{i\lambda t}\, d\lambda = \tfrac{1}{2\pi}\sum_{u=0}^{n-1} X_u \; x \begin{cases} 2\pi, & t=u \\ 0, & t\neq u \end{cases} = \tfrac{1}{2\pi}\, X_t\, 2\pi \;=\; X_t.$$

Thus we have shown (3.1.17). As before, if $X_t$ is real,

(3.1.18) $$J(-\lambda) \;=\; \tfrac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i(-\lambda)t} = \left\{\tfrac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i\lambda t}\right\}^{*} \;=\; J(\lambda)^{*}.$$

**Parseval's Relation**

Now let us consider,

$$\sum_{t=0}^{n-1} |\, X_t\, |^2 = \sum_{t=0}^{n-1} X_t\, X_t^{\,*} \;=\; \sum_{t=0}^{n-1}\sum_{j=0}^{n-1} J_j\, e^{-i\lambda_j t} \sum_{k=0}^{n-1} J_k^{*} e^{i\lambda_k t}$$

We may reorder the summations to obtain

$$\sum_{t=0}^{n-1} |\, X_t\, |^2 = \sum_{j=0}^{n-1} J_j \sum_{k=0}^{n-1} J_k^{*} \sum_{t=0}^{n-1} e^{i(\lambda_k-\lambda_j)t} = \sum_{j=0}^{n-1} J_j \sum_{k=0}^{n-1} J_k^{*}\; x \begin{cases} n, & j=k \bmod(n) \\ 0, & j\neq k \bmod(n) \end{cases}.$$

A minor simplification yields

$$\sum_{t=0}^{n-1} |\, X_t\, |^2 = \sum_{j=0}^{n-1} J_j\, J_j^{\,*}\, n \;=\; n \sum_{j=0}^{n-1} |\; J_j \;|^{\,2}.$$

The equation

(3.1.19) $$\sum_{t=0}^{n-1} |\, X_t \,|^{\,2} \;=\; n \sum_{j=0}^{n-1} |\; J_j \,|^{\,2}$$

is called *Parseval's relation.* Similarly,

$$\int_{-\pi}^{\pi} |\, J(\lambda)\,|^{\,2} d\lambda \;=\; \int_{-\pi}^{\pi} \tfrac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i\lambda t}\; \tfrac{1}{n}\sum_{u=0}^{n-1} X_u^{*}\, e^{i\lambda u}\, d\lambda.$$

Interchanging summations and integration yields

$$\int_{-\pi}^{\pi} |\, J(\lambda)\,|^{\,2} d\lambda \;=\; \tfrac{1}{n^2} \sum_{t=0}^{n-1}\sum_{u=0}^{n-1} X_t\, X_u^{*} \int_{-\pi}^{\pi} e^{i\lambda(u-t)}\, d\lambda.$$

Using (3.1.16)

$$\int_{-\pi}^{\pi} |\, J(\lambda)\,|^{\,2} d\lambda \;=\; \tfrac{1}{n^2} \sum_{t=0}^{n-1}\sum_{u=0}^{n-1} X_t\, X_u^{*}\, x \begin{cases} 2\pi, & t=u \\ 0, & \text{otherwise.} \end{cases}$$

From this it follows that

(3.1.20) $$\int_{-\pi}^{\pi} |\, J(\lambda)\,|^{\,2} d\lambda \;=\; \tfrac{2\pi}{n^2} \sum_{t=0}^{n-1} X_t X_t^{*} \;=\; \tfrac{2\pi}{n^2} \sum_{t=0}^{n-1} |\; X_t\,|^2 .$$

This is also called *Parseval's relation.* We shall see later that there is yet another form of this relation in continuous time. Basically this relation says that the sum of squares in time domain is equal to the sum of squares in frequency domain. One important implication is that the least squares solutions for a fitting problem may be done in either frequency or time domain depending on which is more convenient.

**Some Examples of Fourier Transforms**

In the following discussion we calculate some specific Fourier transforms. First recall that we had earlier shown

$$\sum_{t=0}^{n-1} \cos(\alpha t) = \cos\left(\frac{\alpha(n-1)}{2}\right) \frac{\sin\left(\frac{\alpha n}{2}\right)}{\sin\left(\frac{\alpha}{2}\right)}.$$

In a similar way

(3.1.21) $$\sum_{t=0}^{n-1} e^{i\alpha t} = e^{\frac{i\alpha(n-1)}{2}}\, \frac{\sin\left(\frac{\alpha n}{2}\right)}{\sin\left(\frac{\alpha}{2}\right)}.$$

We may use (3.1.21) to calculate the Fourier transform in a number of interesting cases.

***Cosine Wave:*** First, let

(3.1.22) $$Xt \;=\; R\cos(\lambda t + \phi) \;=\; R\left\{\frac{e^{\,i(\lambda t+\phi)} + e^{\,i(\lambda t-\phi)}}{2}\right\}.$$

Using (3.1.22) in (3.1.4)

$$J_j \;=\; \tfrac{1}{n} \sum_{t=0}^{n-1} X_t e^{-i\lambda_j t} \;=\; \tfrac{R}{n} \sum_{t=0}^{n-1} \cos(\lambda t + \phi)\, e^{\,-i\lambda_j t}.$$

Using the Eulerian identity for cosine, we have

$$J_j \;=\; \tfrac{R}{2n} \sum_{t=0}^{n-1} e^{\,i\lambda t - i\lambda_j t + i\phi} + \tfrac{R}{2n} \sum_{t=0}^{n-1} e^{-i\lambda t - i\lambda_j t - i\phi}.$$

Collecting terms and simplifying

$$J_j \;=\; \tfrac{R}{2n}\, e^{i\phi} \sum_{t=0}^{n-1} e^{i(\lambda-\lambda_j)t} + \tfrac{R}{2n}\, e^{-i\phi} \sum_{t=0}^{n-1} e^{\,-i(\lambda+\lambda_j)t}.$$

These are each geometric series and may, as before, be written in terms of the Dirichlet kernel. Thus we have

$$J_j = \frac{R}{2n} e^{i\phi} e^{\frac{i(n-1)(\lambda-\lambda_j)}{2}} \frac{\sin\left(\frac{n(\lambda-\lambda_j)}{2}\right)}{\sin\left(\frac{(\lambda-\lambda_j)}{2}\right)} + \frac{R}{2n} e^{-i\phi} e^{\frac{-i(n-1)(\lambda+\lambda_j)}{2}} \frac{\sin\left(\frac{n(\lambda+\lambda_j)}{2}\right)}{\sin\left(\frac{(\lambda+\lambda_j)}{2}\right)}$$

which may be rewritten as

$$J_j = \frac{R}{2} e^{i\phi} e^{\frac{i(n-1)(\lambda-\lambda_j)}{2}} D_n(\lambda-\lambda_j) + \frac{R}{2} e^{-i\phi} e^{\frac{-i(n-1)(\lambda+\lambda_j)}{2}} D_n(\lambda+\lambda_j).$$

Or finally,

(3.1.23) $$J_j = \frac{R}{2} e^{\frac{-i(n-1)\lambda_j}{2}} \left\{ e^{\frac{i((n-1)\lambda+2\phi)}{2}} D_n(\lambda-\lambda_j) + e^{\frac{-i((n-1)\lambda+2\phi)}{2}} D_n(\lambda+\lambda_j) \right\}.$$

This illustrates the *leakage phenomena.* The leakage phenomena refers to the fact that when one takes a Fourier transform of a single sinusoid, i.e. all energy concentrated at a single frequency, the Fourier transform suggests that energy has leaked out to neighboring frequencies.

***A Step Function:*** Let

(3.1.24) $$X_t = \begin{cases} 1 & 0 \le t < m \\ & \\ 0 & m \le t < n \end{cases}.$$

In this case

$$Jj = \frac{1}{n} \sum_{t=0}^{m-1} e^{i\lambda_j t} = \frac{1}{n} e^{\frac{-i\lambda_j(m-1)}{2}} \frac{\sin\left(\frac{(m-1)\lambda_j}{2}\right)}{\sin\left(\frac{\lambda_j}{2}\right)}.$$

Rewriting in terms of the Dirichlet kernel

(3.1.25) $$J_j = \frac{m}{n} e^{\frac{-i\lambda_j(m-1)}{2}} D_m(\lambda_j).$$

Thus

(3.1.26) $$R^2(\lambda_j) = J_j J_j^* = (\frac{m}{n})^2 \mid D_m(\lambda_j) \mid^2.$$

***The Linear Process*:** Now let

(3.1.27) $$X_t = t - \frac{n-1}{2},\ t = 0, \ldots, n-1 \text{ (chosen to give a mean of 0.)}$$

Let us first recall for a geometric series

****

(3.1.28) $$\sum_{t=0}^{n} a^t = 1 + \sum_{t=1}^{n} a^t = \frac{1-a^{n+1}}{1-a}$$

Taking the derivative of (3.1.28) with respect to a,

$$\frac{d\left(\sum_{t=0}^{n} a^t\right)}{d\,a} = \sum_{t=1}^{n} t\, a^{t-1} = \frac{d\left(\frac{1-a^{n+1}}{1-a}\right)}{da}$$

so that

$$\sum_{t=1}^{n} t\, a^{t-1} = \frac{(1-a)(-(n+1)\,a^n)-(1-a^{n+1})(-1)}{(1-a)^2}$$

which is

$$\sum_{t=1}^{n} t\, a^{t-1} = -\frac{(n+1)(1-a)a^n-(1-a^{n+1})}{(1-a)^2}.$$

Thus

$$\sum_{t=0}^{n} ta^t = a\sum_{t=0}^{n} ta^{t-1} = -\frac{(n+1)(1-a)a^{n+1}-a(1-a^{n+1})}{(1-a)^2}.$$

Finally replacing n by n – 1,

(3.1.29) $$\sum_{t=0}^{n-1} ta^t = -\frac{n(1-a)\,a^n-a(1-a^n)}{(1-a)^2}.$$

Thus

$$J_j = \frac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i\lambda_j t} = \frac{1}{n}\sum_{t=0}^{n-1} t\, e^{-i\lambda_j t} - \frac{n-1}{2n}\sum_{t=0}^{n-1} e^{-i\lambda_j t}.$$

****

$$J_j = \frac{(1-e^{-i\lambda_j})e^{-i\lambda_j n}}{(1-e^{-i\lambda_j})^2} - \frac{e^{-i\lambda_j}\left(1-e^{-i\lambda_j n}\right)}{n\left(1-e^{-i\lambda_j}\right)^2}$$

$$- \frac{n-1}{2n}\left\{e^{-i\lambda_j\frac{(n-1)}{2}}\,\frac{\sin\left((n-1)\lambda_j/2\right)}{\sin(\lambda_j/2)}\right\}.$$

Hence we may simplify to get

(3.1.30) $$J_j = \left(1-e^{-\lambda_j}\right)^{-1} e^{-i\lambda_j n} - \frac{1}{n}\, e^{-i\lambda_j}\left(1-e^{-i\lambda_j n}\right)\left(1-e^{-i\lambda_j}\right)^{-2}$$

$$- \frac{n-1}{2}e^{-i\lambda_j\frac{(n-1)}{2}}D_n(\lambda_j).$$

**Other Properties of the Fourier Transform**

***Time Shift:*** Consider

(3.1.31) $$Y_t = X_{t+h}, \quad t = 0, \ldots, n-1.$$

Then we may let

$$J_{y,j} = \frac{1}{n}\sum_t Y_t\, e^{-i\lambda_j t} = \frac{1}{n}\sum_t X_{t+h}\, e^{-i\lambda_j t}.$$

Now we change the variable of summation by letting u = t + h so that t = u – h to obtain

(3.1.32) $$J_{y,j} = \frac{1}{n}\sum_u X_u\, e^{-i\lambda_j(u-h)} = \frac{e^{i\lambda_j h}}{n}\sum_u X_u\, e^{-i\lambda_j u} = e^{i\lambda_j h}J_{x,j}.$$

***Symmetry:*** Let

(3.1.33) $$X_t = X_{n-t}, t = 0, \ldots, n.$$

Thus we have that

$$J_j = \tfrac{1}{n+1} \sum_{t=0}^{n} X_t \, e^{-i\lambda_j t} = \tfrac{1}{n+1} \sum_{t=0}^{n} X_{n-t} \, e^{-i\lambda_j t}.$$

Again we change variables by letting $u = n - t$ so that $t = n - u$. We thus obtain

$$J_j = \tfrac{1}{n+1} \sum_{u=0}^{n} X_u \, e^{-i\lambda_j (n-u)} = \tfrac{1}{n+1} \sum_{u=0}^{n} X_u \, e^{i\lambda_j u} e^{i\lambda_j n}.$$

Recall that $\lambda_j = 2\pi(j/n)$ so that $\lambda_j n = 2\pi j$. Thus $e^{i\lambda_j n} = 1$ and

(3.1.34) $$J_j = \tfrac{1}{n+1} \sum_{u=0}^{n} X_u \, e^{+i\lambda_j u}.$$

If $X_t$ is a real-valued time series, we may take the complex conjugate of (3.1.34) to obtain

(3.1.35) $$J_j^* = \tfrac{1}{n+1} \sum_{u=0}^{n} X_u^* \left(e^{+i\lambda_j u}\right)^* = \tfrac{1}{n+1} \sum_{u=0}^{n} X_u \, e^{-i\lambda_j u} = J_j.$$

**Roughness Measures:**

Let us finally consider

(3.1.36) $$\sum_{t=1}^{n-1} (X_t - X_{t-1})^2$$

as a measure of roughness of the process. This is essentially the variation of the process. Now,

$$X_t - X_{t-1} = \sum_{j=0}^{n-1} J_j \left\{ e^{i\lambda_j t} - e^{i\lambda_j (t-1)} \right\} = \sum_{j=0}^{n-1} J_j \, e^{i\lambda_j t} \left(1 - e^{-i\lambda_j}\right).$$

Thus, we may write

****

$$(X_t - X_{t-1})^2 = \sum_{j=0}^{n-1} J_j \, e^{i\lambda_j t} \left(1 - e^{-i\lambda_j}\right) \sum_{k=0}^{n-1} J_k^* \, e^{-i\lambda_k t} (1 - e^{+i\lambda_k})$$

so that

(3.1.37) $$\sum_{t=1}^{n-1} (X_t - X_{t-1})^2 = \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} J_j \, J_k^* \left(1 - e^{-i\lambda_j}\right)\left(1 - e^{i\lambda_k}\right) \sum_{t=1}^{n-1} e^{i(\lambda_j - \lambda_k)t}$$

which by (3.1.3) yields

(3.1.38) $$\sum_{t=1}^{n-1} (X_t - X_{t-1})^2 = \sum_{j=0}^{n-1} |J_j|^2 \, |1 - e^{i\lambda_j}|^2.$$

But

$$\left(1 - e^{i\lambda_j}\right)\left(1 - e^{-i\lambda_j}\right) = 2 - \left(e^{i\lambda_j} + e^{-i\lambda_j}\right) = 2 - 2\cos(\lambda_j) = 2(1 - \cos(\lambda_j)).$$

Thus

$$\sum_{t=1}^{n-1} (X_t - X_{t-1})^2 = 2 \sum_{j=0}^{n-1} |J_j|^2 (1 - \cos(\lambda_j)).$$

But by elementary trigonometry, $1 - \cos(\lambda_j) = 2 \sin^2(\lambda_j/2)$, so that

(3.1.39) $$\sum_{t=1}^{n-1} (X_t - X_{t-1})^2 = 4 \sum_{j=0}^{n-1} |J_j|^2 \sin^2(\lambda_j/2).$$

We may use Parseval's relation, $\sum_{t=0}^{n-1} X_t^2 = n \sum_{j=0}^{n-1} |J_j|^2$, to scale equation (3.1.39) to obtain

(3.1.40) $$\sum_{t=1}^{n-1} (X_t - X_{t-1})^2 \Big/ 2 \sum_{t=0}^{n-1} X_t^2 = \frac{4 \sum_{j=0}^{n-1} \sin^2(\lambda_j/2) |J_j|^2}{n \sum_{j=0}^{n-1} |J_j|^2}.$$

This equation may be used as a measure of frequency content. If low frequencies predominate, $X_t$ is close to $X_{t-1}$ and the ratio is small. If high frequencies are plentiful, the ratio will be large.

# Chapter 4
# Fast Fourier Transforms

Let us return to our consideration of the discrete Fourier transform (DFT) of $X_t$. Let $\lambda_j = 2\pi(j/n)$ as before for $j = 0, 1, \dots, n-1$. Recall that

$$n J_j = \sum_{t=0}^{n-1} X_t \, e^{-i\lambda_j t}, \quad j = 0, 1, \dots, n-1. \tag{4.1.1}$$

In this computation, there are $n-1$ (complex) multiplies (one for each $j = 1, \dots, n-1$; $j = 0$ implies $e^0 = 1$ so that no multiply is involved). There are also $(n-1)$ additions (adding n terms means $n-1$ additions). Since there are a total of n equations, to compute the full Fourier transform, there are $n(n-1)$ complex additions and $n(n-1)$ complex multiplies.

Now if we notice that $e^{-i\lambda_j t} = e^{-2\pi i \frac{jt}{n}}$, t and $j = 0, 1, \dots, n-1$ and observe that $e^{2\pi i k} = 1$ for all integer k, there are only n distinct values of $e^{-i\lambda_j t}$. The idea of the fast Fourier transform (FFT) is basically to collect the coefficients of the n distinct $e^{-i\lambda_j t}$ depending on $j \times t$ mod (n) and do only n multiplications. The simplest case is the case that n factors say $n = n_1 \times n_2$. If n is prime there is no FFT. The generalized situation is $n = n_1 \times n_2 \times \cdots \times n_k$. The most useful case is the special case that $n = 2^k$, i. e. all factors are 2. They need not be, however; the factors only need be prime. (Actually, primes are required for most efficient implementation. The FFT can be done with non-primes, but a more efficient prime-based FFT can always be done.) Let us illustrate with the case that $n = n_1 \times n_2$.

Write

$$t = t_1 n_2 + t_2, \quad 0 \le t_1 < n_1, \quad 0 \le t_2 < n_2$$

to generate a rectangular array, e.g. $n = 55 = 5 \text{ x } 11$.

$$\begin{array}{llll} X_0 & X_1 & \dots & X_{10} \\ X_{11} & X_{12} & \dots & X_{21} \\ \vdots & \vdots & \vdots & \vdots \\ X_{44} & X_{45} & \dots & X_{54} \end{array}$$

or, in general

$$\begin{array}{llll} X_0 & X_1 & \dots & X_{n_2-1} \\ X_{n_2} & X_{n_2+1} & \dots & X_{2n_2-1} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n-n_2} & X_{n-n_2+1} & \dots & X_{n_1 n_2-1}. \end{array}$$

Let $Y(t_1, t_2) = X_{t_1 n_2 + t_2}$ and let $j = j_2 n_1 + j_1$, $0 \le j_2 < n_2$, $0 \le j_1 < n_1$. For integer $\alpha$, let $w_\alpha = e^{-2\pi i/\alpha}$. Then

$$n J_j = \sum_{t=0}^{n-1} X_t \, e^{-\frac{2\pi i j t}{n}} = \sum_{t_1=0}^{n_1-1} \sum_{t_2=0}^{n_2-1} X_{t_1 n_2 + t_2} \, w_n^{(j_2 n_1 + j_1)(t_1 n_2 + t_2)} \tag{4.1.2}$$

or equivalently

$$n J_j = \sum_{t_1=0}^{n_1-1} \sum_{t_2=0}^{n_2-1} Y(t_1, t_2) \, w_n^{\,t_1 j_2 n_1 n_2 + j_1 t_1 n_2 + j_2 t_2 n_1 + j_1 t_2}. \tag{4.1.3}$$

But $w_n^{t_1 j_2 n_1 n_2} = \left\{e^{2\pi i}\right\}^{t_2 j_2} = 1^{t_1 j_2} = 1$. Thus (4.1.3) simplifies to

(4.1.4) $$n\,J_j = \sum_{t_1=0}^{n_1-1} \sum_{t_2=0}^{n_2-1} Y(t_1, t_2)\, w_n^{\,j_1 t_1 n_2}\, w_n^{\,j_2 t_2 n_1}\, w_n^{j_1 t_2}.$$

Also,

$$w_n^{n_1} = e^{-\frac{2\pi i}{n} n_1} = e^{-\frac{2\pi i}{n_2}} = w_{n_2}$$

and, similarly,

$$w_n^{n_2} = e^{-\frac{2\pi i}{n} n_2} = e^{-\frac{2\pi i}{n_1}} = w_{n_1}.$$

Reversing the order of summation in (4.1.4),

*****

(4.1.5) $$n\,J_j = \sum_{t_2=0}^{n_2-1} w_{n_2}^{j_2 t_2}\, w_n^{j_1 t_2} \sum_{t_1=0}^{n_1-1} Y(t_1, t_2)\, w_{n_1}^{j_1 t_1}.$$

Consider the last summation in (4.1.5)

(4.1.6) $$Z(j_1, t_2) = \sum_{t_1=0}^{n_1-1} Y(t_1, t_2)\, w_{n_1}^{j_1 t_1}.$$

This is a column by column transform of the original data. In this transform, there are $n_1 - 1$ multiplies and $n_1 - 1$ adds. Now

(4.1.7) $$n\,J_j = \sum_{t_2=0}^{n_2-1} \left\{ w_n^{j_1 t_2}\, Z(j_1, t_2) \right\} w_{n_2}^{j_2 t_2}.$$

For a fixed j there are $n_1 - 1$ adds in equation (4.1.6) and $n_2 - 1$ adds in equation (4.1.7). Thus there are $n_1 + n_2 - 2$ adds for each j and total of $n(n_1 + n_2 - 2) = n_1 n_2 (n_1 + n_2 - 2)$. For multiplies, we have $n_1 - 1$ multiplies in (4.1.6). In addition in (4.1.7), we have $n_2 - 1$ multiplies for a fixed j, so that we have $n_1 + n_2 - 2$ for each fixed j. Since we have n j 's we have $n(n_1 + n_2 - 2)$ multiplies. The multiplication of the twiddle factor requires n additional multiplies ($j_1 = 0, \ldots, n_1 - 1$, $t_2 = 0, \ldots, n_2$). Thus the total multiplies for the factored form is

$$n(n_1 + n_2 - 2) + n = n_1 n_2 (n_1 n_2 - 1)$$

*Example:* Let $n = 50 = 5 \times 10$ and consider the full transform DFT. Here $50(50 - 1) = 50 \times 49 = 2450$, so that we have 2450 adds and 2450 multiplies. For the FFT, we have $50(15 - 2) = 50 \times 13 = 650$ adds and $50\,(15 - 1) = 50 \times 14 = 700$ multiplies. This gives a FFT/DFT ratio of 650/2450 = .2653 for adds and 700/2450 = .2857 for multiplies.

*Example*: Consider this time $n = 49 = 7 \times 7$ in the full DFT. Here $49(49 - 1) = 49 \times 48 = 2352$. Thus we have 2352 adds and 2352 multiplies. Correspondingly for the FFT, we have $49\,(7 + 7 - 2) = 49 \times 12 = 588$ adds and $49\,(14 - 1) = 49 \times 13 = 637$ multiplies. Thus the FFT/DFT ratio is 588/2352 = .25 for adds and 637/2352 = .2708 for multiplies. We minimize these ratios when n, the number of observations, is fully factored.

As we indicated before, if $n = n_1 \cdots n_k$, we may arrange the observations over a k – dimensional table. In this case the number of adds (and multiplies) is $n_1 \cdots n_k \left(\sum_{j=1}^{k} n_j - c\right)$ where c depends on exact details. But, in

any case, $n_1 \cdots n_k(\sum_{j=1}^{k} n_j)$ is an upper bound on the number of adds and the number of multiplies. Thus we may consider further examples.

*Example:* Now let $n = 50 = 5 \times 5 \times 2$ so that n is fully factored. Then 50(12) is an upper-bound and actually in this case 50 (5 + 5 + 2 − 3) is the full correct computation for the number of adds. Notice that 50 × 12 = 600, but 50 (9) = 450. Thus the prime factorization results in a savings of 200 adds and an efficiency ratio of 450/2450 = .1837.

Frequently, we consider $n = 2^k$, i.e. $\log_2 n = k$. Then an upper bound is $n(n_1 + \cdots + n_k) = n(2 + \cdots + 2) = 2kn = 2n \times \log_2 n$ for the FFT which is compared to $n^2$ for the standard DFT. The FFT/DFT efficiency ratio is thus $(2\log_2 n)/n$ when the sample size is a power of two.

*Example:* Consider now n = 128, so that k = 7 and the FFT/DFT efficiency ratio is 14/64 = .1875. Thus an FFT for 64 terms would take 2 × 64 × 6 = 768 adds and multiplies. Recall for n = 50 = 10 × 5, we had 700 multiplies and 650 adds. One can conclude from this example that not much is lost in absolute operation count by using FFT with excess terms since we have essentially the same operation counts for an FFT of 128 terms and a DFT of only 50 terms.

*Example:* Consider an even more extreme example with n = 1024, so that k = 10 making the FFT/DFT efficiency ratio to be .019 = $\frac{20}{1024}$, i.e. 20 × 1024 = 20480 compared to 1024(1023) = 1,047,552. Suppose now we have $X_0, \ldots, X_{n-1}$ with $n < 2^k = n'$. Let $X'_j = X_j$, j = 0, 1,..., n − 1 and let $X'_j = 0$, j = n,..., n′ − 1 so that we fill the data vector out with zeros. Then

$$J'_j = \frac{1}{n'} \sum_{t=0}^{n'-1} X'_t \, e^{-i\lambda'_j t}.$$

Here we will let $\lambda'_j = 2\pi j/n'$. Then $J'_j = \frac{1}{n'} \sum_{t=0}^{n-1} X_t \exp(-i\lambda'_j t)$

$$= \frac{n}{n'} \times \frac{1}{n} \sum_{t=0}^{n-1} X_t \exp(-i\lambda'_j t) = \frac{n}{n'} J(\lambda'_j).$$

Thus $J(\lambda_j) = \frac{n'}{n} J'_j$. The FFT of the extended series is the Fourier transform of the original series evaluated at $\lambda'_j$ instead of $\lambda_j$. Of course,

$$J_j = J(\lambda_j) \neq J(\lambda'_j), \text{ but since } n' > n, \lambda'_j = 2\pi j/n' < 2\pi j/n = \lambda_j.$$

Thus we evaluate $J(\lambda)$ on a denser grid when we fill out the FFT with zeros.

*Example: (Worst case).* Let n = 1025 so that n = 1024 too small. Extend the sample size to $n' = 2048$ with 1023 zeros filled in. Then the number of FFT multiplies for $n' = 2048 = 2^{11}$ is $2 \times 11 \times 2048 = 45{,}056$. The number of FFT multiplies for $n = 1024 = 2^{10}$ is $2 \times 10 \times 1024 = 20{,}480$. However, the number of DFT multiplies for n = 1025 is $1025 \times 1024 = 1{,}049{,}600$ so that even in the worst case scenario the FFT is dramatically better than the DFT. Consider for a moment the corresponding evaluation grids. For the n = 2048 FFT case $\lambda_j = 2\pi j/2048$, j = 0,...,2047 so that the evaluation points are 0, .0003069, .00614, .00921, .01227, .... The evaluation grid for the n=1025 DFT case is computed based on $\lambda_j = 2\pi j/1025$, j = 0,...,1025. This gives a grid of evaluation points 0, .00612, .01226, ... which is roughly twice the spacing of the more desirable FFT grid.

# Chapter 5
# Leakage and Aliasing

As we have already seen, harmonic analysis can lead to substantial insight into the frequency structure of a time series. Harmonic methods based on Fourier transforms are not, however, universal panaceas for understanding the frequency content of a time series. Problems exist with so-called leakage and aliasing and with lack of controllability of the variance of raw Fourier transforms. In this chapter we shall look into these problems in somewhat more detail.

### 5.1 Leakage

Leakage refers to the fact that if we take the Fourier transform of a pure sinusoidal function, we will get non-zero amplitude for nearby frequencies. To see this, let

$$X_t = \exp(i\lambda_0 t),$$

the complex sinusoid with unit amplitude and zero phase, then

$$J(\lambda) = \frac{1}{n}\sum_{t=0}^{n-1} X_t\, e^{-i\lambda t} \tag{5.1.1}$$

so that

$$J(\lambda) = \frac{1}{n}\sum_{t=0}^{n-1} e^{i(\lambda_0-\lambda)t}. \tag{5.1.2}$$

Using results from Chapter 3

$$J(\lambda) = e^{\frac{i(n-1)(\lambda_0-\lambda)}{2}} \times \frac{\sin(n(\lambda_0-\lambda)/2)}{n\sin((\lambda_0-\lambda)/2)}. \tag{5.1.3}$$

Taking the square modulus of the expression in (5.1.3), we have

$$R^2(\lambda) = J(\lambda)\,J(\lambda)^* = \left(\frac{\sin^2(n(\lambda_0-\lambda)/2)}{n^2\sin^2((\lambda_0-\lambda)/2)}\right). \tag{5.1.4}$$

It is easy to see from (5.1.4) that there is spectral energy near $\lambda_0$. $R^2(\lambda)$ is plotted in Figure ****. Let us further consider K defined by

$$K(\lambda) = J(\lambda)\, e^{\frac{i(n-1)\lambda}{2}} \tag{5.1.5}$$

which may be rewritten

$$K(\lambda) = e^{\frac{i(n-1)\lambda_0}{2}} \times \frac{\sin(n(\lambda_0-\lambda)/2)}{n\sin((\lambda_0-\lambda)/2)}. \tag{5.1.6}$$

We may shift the argument of k by $2\pi/n$ to obtain

$$K(\lambda - \tfrac{2\pi}{n}) = e^{\frac{i(n-1)\lambda_0}{2}} \times \frac{\sin\left(n\left(\lambda_0-\left(\lambda-\frac{2\pi}{n}\right)\right)/2\right)}{n\sin\left(\left(\lambda_0-\left(\lambda-\frac{2\pi}{n}\right)\right)/2\right)}. \tag{5.1.7}$$

We may simplify to obtain

(5.1.8) $$K(\lambda - \tfrac{2\pi}{n}) \;=\; e^{\frac{i(n-1)\lambda_0}{2}} \times \frac{\sin\left(n(\lambda_0-\lambda)/2+\pi\right)}{n\sin\left(\left(\frac{\lambda_0-\lambda}{2}\right)+\frac{\pi}{n}\right)},$$

so that approximately

(5.1.9) $$K(\lambda - \tfrac{2\pi}{n}) \;\doteq\; -\; e^{\frac{i(n-1)\lambda_0}{2}} \times \frac{\sin\left(n(\lambda_0-\lambda)/2\right)}{n\sin\left((\lambda_0-\lambda)/2\right)}.$$

Similarly, we can show that

(5.1.10) $$K\left(\lambda + \tfrac{2\pi}{n}\right) \;\doteq\; -\; e^{\frac{i(n-1)\lambda_0}{2}} \times \frac{\sin\left(n(\lambda_0-\lambda)/2\right)}{n\sin\left((\lambda_0-\lambda)/2\right)}.$$

Let us define $K_h(\lambda)$ by

(5.1.11) $$K_h(\lambda) \;=\; \tfrac{1}{4}\,K\left(\lambda - \tfrac{2\pi}{n}\right) \;+\; \tfrac{1}{2}\,K(\lambda) \;+\; \tfrac{1}{4}\,K\,(\lambda + \tfrac{2\pi}{n}).$$

Notice for $\lambda \neq \lambda_0$, $K_h(\lambda) \;\doteq\; 0$ for n sufficiently large for the Dirichlet approximation to hold. In reality we are interested in $J(\lambda)$, not $K(\lambda)$, so we transform back by taking $J_h$ to be defined by

(5.1.12) $$K_h(\lambda) \;=\; J_h(\lambda)\, e^{\frac{i(n-1)\lambda}{2}}$$

or equivalently

(5.1.13) $$J_h(\lambda) \;=\; K_h(\lambda)\, e^{\frac{-i(n-1)\lambda}{2}}.$$

This weighted sum is called the *Hanning Window* and by it we introduce the idea that we can reduce leakage by taking weighted sums. A little algebraic manipulation will lend some further insight. First, since (5.1.11) gives us

$$K_h(\lambda) = \tfrac{1}{4}\,K(\lambda - \tfrac{2\pi}{n}) + \tfrac{1}{2}\,K(\lambda) + \tfrac{1}{4}\,K(\lambda + \tfrac{2\pi}{n}),$$

we have

$$K_h(\lambda) = J(\lambda - \tfrac{2\pi}{n}) \times \tfrac{1}{4} \times\; e^{\frac{i(n-1)(\lambda-\frac{2\pi}{n})}{2}} + J(\lambda) \times \tfrac{1}{2} \times\; e^{\frac{i(n-1)\lambda}{2}} + J(\lambda + \tfrac{2\pi}{n}) \times \tfrac{1}{4} \times\; e^{\frac{i(n-1)(\lambda+\frac{2\pi}{n})}{2}}.$$

Substituting for the $J(\lambda)$, we have

(5.1.14)
$$K_h(\lambda) = \left(\tfrac{1}{n}\sum_{t=0}^{n-1} X_t\; e^{-i\lambda t}\; e^{\frac{i(n-1)\lambda}{2}}\right) \times \left(\tfrac{1}{4}e^{\frac{-i2\pi t}{n}}e^{-i\frac{2\pi}{n}(\frac{n-1}{2})} + \tfrac{1}{2} + \tfrac{1}{4}e^{\frac{i2\pi t}{n}}\; e^{i\frac{2\pi}{n}(\frac{n-1}{2})}\right).$$

Simplifying we have

(5.1.15) $$K_h(\lambda) = \frac{1}{n}\, e^{\frac{i(n-1)\lambda}{2}} \sum_{t=0}^{n-1} X_t\, e^{-i\lambda t} \left\{ \frac{1}{2} + \frac{1}{2}\cos\left( \frac{2\pi}{n}\left( t + \frac{n-1}{2} \right) \right) \right\}.$$

Thus, we may write

(5.1.16) $$J_h(\lambda) \;=\; \frac{1}{n} \sum_{t=0}^{n-1} X_t \left\{ \frac{1-\cos(2\pi(t+\frac{1}{2})/n}{2} \right\} e^{-i\lambda t}.$$

We may define a weighted data set by

(5.1.17) $$Y_t = X_t \left\{ \frac{1-\cos(2\pi(t+\frac{1}{2})/n)}{2} \right\}.$$

Thus we are led to the idea of weighting the data to reduce this bias introduced by leakage. We have seen that we can do this in the frequency domain by taking the weighted sums of Fourier transforms

$$K_h(\lambda) \;=\; \frac{1}{4} K\,(\lambda - \tfrac{2\pi}{n}) + \frac{1}{2} K\,(\lambda) + \frac{1}{4} K\,(\lambda + \tfrac{2\pi}{n})$$

or in the time domain by weighting the data

$$Y_t \;=\; X_t \left\{ \frac{1-\cos(2\pi(t+\frac{1}{2})/n)}{2} \right\}$$

The latter is called the *data window*, the former is usually called the *spectral window*. We shall investigate windowing in much more detail later on.

### 5.2 Convolutions

One interesting little sidelight to note concerns convolutions. Let $Z_t$ be defined by

(5.2.1) $$Z_t = \sum_{u=0}^{n-1} X_u\, Y_{t-u}, \quad t = 0\,,1,\ldots,n-1.$$

Then,

(5.2.2) $$J_z(\lambda) = \sum_{t=0}^{n-1} Z_t\, e^{-i\lambda t}.$$

Writing out $Z_t$, we have

(5.2.3) $$J_z(\lambda) = \sum_{t=0}^{n-1} \sum_{u=0}^{n-1} X_u\, Y_{t-u}\, e^{-i\lambda t}.$$

Interchanging the order of summation

(5.2.4) $$J_z(\lambda) = \sum_{u=0}^{n-1} X_u \sum_{t=0}^{n-1} Y_{t-u}\, e^{-i\lambda t}.$$

Let $v = t - u$, so that $v + u = t$. Then we have

(5.2.5) $$J_z(\lambda) = \sum_{u=0}^{n-1} X_u \sum_{v=0}^{n-1} Y_v \, e^{-i\lambda v} \, e^{-i\lambda u}.$$

We can separate variables to obtain

(5.2.6) $$J_z(\lambda) = \sum_{u=0}^{n-1} X_u \, e^{-i\lambda u} \sum_{v=0}^{n-1} Y_v \, e^{-i\lambda v} = J_x(\lambda) \times J_y(\lambda).$$

Thus the Fourier transform of a convolution of two time series is the product of the Fourier transforms of the two time series and, of course, vice versa. The leakage phenomena can be understood more easily in terms of convolutions. Leakage really arises because an infinite set of data, namely $\ldots\ X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$, is multiplied by a finite function, to be precise the function

(5.2.7) $$I_{[0,n]}(j) = \begin{cases} 1 & \text{if } j \in [0, n-1] \\ 0 & \text{if } j \notin [0, n-1]. \end{cases}$$

The calculation we have just performed shows us is that we can reduce leakage bias by replacing $I_{[0,n]}(j)$ by a more tapered window, say

(5.2.8) $$H(j) = \left\{ \frac{1 - \cos(2\pi(j+\frac{1}{2})/n)}{2} \right\}.$$

Of course, this is not the only window we could use, but it is a computationally simple one. We shall use this smoothing window idea again in later discussions.

### 5.3 Aliasing

We have just seen that one problem in dealing with DFTs involves frequency smearing. That is, the DFT of a single frequency signal yields a frequency spectrum which spreads some energy of the single frequency signal over frequencies adjacent to that of the signal. Another interesting phenomena involves aliasing. An alias is another name for something. In the frequency analysis of a continuous time signal we shall see that potentially many frequencies of a signal will appear in the same place, i.e. many frequencies are labeled with the same frequency designation. This is the origin of the name, alias.

Let us begin by considering a single frequency process

(5.3.1) $$X_t = \cos(\lambda t)$$

which is sampled at time increments $\Delta t$, so that $X_j = \cos(\lambda j \Delta t)$. As we vary $\lambda$ toward zero the cosine wave becomes lower in frequency and the wave length, L, increases. However, as $\lambda$ gets larger, specifically as $\lambda \to \frac{\pi}{\Delta t}$

(5.3.2) $$X_j \to \cos\left(\frac{\pi}{\Delta t} j \Delta t\right) = \cos(j\pi) = \begin{cases} -1 \text{ when j is odd} \\ 1 \text{ when j is even} \end{cases}.$$

For $\frac{\pi}{\Delta t} < \lambda < \frac{2\pi}{\Delta t}$, let $\lambda' = \frac{2\pi}{\Delta t} - \lambda$, so that $\lambda = \frac{2\pi}{\Delta t} - \lambda'$. Then

(5.3.3) $$X_j = \cos(\lambda j \Delta t).$$

Substituting for $\lambda$,

(5.3.4) $$X_j = \cos\left\{\left(2\pi\Delta t - \lambda'\right)j\Delta t\right\}$$

which may be simplified to

(5.3.5) $$X_j = \cos(2\pi j - \lambda' j\Delta t) = \cos(-\lambda' j\Delta t) = \cos(\lambda' j\Delta t).$$

That is, based on this sampling rate, we can't distinguish between $\lambda$ and $\lambda'$. Next consider $\frac{2\pi}{\Delta t} < \lambda < \frac{3\pi}{\Delta t}$, let $\lambda'' = \frac{2\pi}{\Delta t} + \lambda$, so that $\lambda'' - \frac{2\pi}{\Delta t} = \lambda$. By a similar computation,

$$X_j = \cos(\lambda j\Delta t) = \cos\left\{\left(\lambda'' - \tfrac{2\pi}{\Delta t}\right)j\Delta t\right\} = \cos\left\{\lambda'' j\Delta t - 2\pi j\right\} = \cos\left(\lambda'' j\Delta t\right).$$

Graphically, frequencies are folded back upon each other. See Figures **** Because of this folding pattern, (and, of course, that Nyquist first articulated this idea), the frequency, $\pi/\Delta t$, is called the *Nyquist folding frequency*. For a sampled time series frequencies higher that the Nyquist frequency are indistinguishable from (alias as) lower frequencies. This process is called *aliasing*.

# Chapter 6
# Random Process Models

## 6.1 Introduction to Random Processes

To this point we have said nothing of the statistical structure of the process $X_t$, especially with respect to the DFT of $X_t$. The DFT is of course itself a random process if $X_t$ is a random process. The implication is that a different realization of $X_t$ yields a different realization of $J(\lambda)$. In order to make the DFT a statistically meaningful diagnostic tool, we must investigate the stochastic structure of $X_t$ a bit more thoroughly. Basically, the use of the DFT straight away on the data implies we believe $X_t$ is essentially deterministic.

We now want to regard $X_t$ as a stochastic process. That is to say we let $\Omega$ be a sample space and P a probability on that sample space satisfying the axioms given below.

Axiom 1. For $A \subseteq \Omega$, $0 \leq P(A) \leq 1$.

Axiom 2. For $A_i \subseteq \Omega$, $A_i$ disjoint, $\sum_i P(A_i) = P(\bigcup_i A_i)$.

Axiom 3. $P(\Omega) = 1$

In this section it is often convenient to regard a random process, X, as complex valued rather than real valued. Of course, since all real-valued processes are trivially complex-valued, this is no real loss of generality. We shall define $X_t(\omega) = X(t, \omega)$ to be a random process if

$$X(t, \cdot) : \Omega \rightarrow \mathbb{R}, \text{ the real numbers (or } \mathbb{C}, \text{ the complex numbers).}$$

That is to say, for t fixed, $X(t, \cdot)$ is a mapping of the sample space into the real line (the complex plane). For each fixed t, $X(t, \omega)$ is an ordinary random variable. Usually we write $X_t(\cdot)$ for $X(t, \cdot)$. It is common to suppress the explicit dependence on $\omega$, although, of course, it is always there implicitly.

The variable, t, is usually thought of as belonging to an index set, $\mathcal{T}$, and usually is thought of as time. As we had earlier discussed in Chapter 1,

$$F_{x_{t_1},\ldots,x_{t_n}}(u_1,\ldots,u_n) = P(X_{t_1} \leq u_1, \ldots, X_{t_n} \leq u_n) \tag{6.1.1}$$

is the finite dimensional distribution. $X_t$ is *strictly stationary* if

$$F_{x_{t_1},\ldots,x_{t_n}}(u_1,\ldots,u_n) = F_{x_{t_{1+\tau}},\ldots,x_{t_n+\tau}}(u_1,\ldots,u_n) \tag{6.1.2}$$

for every n, $t_1, \ldots, t_n$, $u_1, \ldots, u_n$ and $\tau$. Again recall that if $\mu_t =$ a constant, usually 0 and cov $(X_t, X_{t+\tau}) = \gamma_\tau$, then we say $X_t$ is *weakly stationary*.

Example: Let $X_t = \sum_{n=1}^{k} \{A_n \cos(n\lambda t) + B_n \sin(n\lambda t)\}$. Here $A_n$ and $B_n$ are independent normal random variables with mean $\mu = 0$ and variance $= \sigma^2$, written as $\mathcal{N}(0, \sigma^2)$.

Taking expectations we have

$$E(X_t) = \sum_{n=1}^{k} \{E(A_n)\cos(n\lambda t) + E(B_n)\sin(n\lambda t)]\} = 0.$$

Consider now the product term and take expectation to get

$$E\{X_{t_1}X_{t_2}\} = E\left\{\sum_{n=1}^{k}\sum_{m=1}^{k}\left(A_n \cos(n\lambda t_1) + B_n \sin(n\lambda t_2)\right)\left(A_m \cos(m\lambda t_2) + B_m \sin(m\lambda t_2)\right)\right\}.$$

Expanding we obtain,

$$E\{X_{t_1}X_{t_2}\} = \sum_{n=1}^{k}\left\{E(A_n^2)\cos(n\lambda t_1)\cos(n\lambda t_2) + E(B_n^2)\sin(n\lambda t_1)\sin(n\lambda t_2)\right\},$$

so that

$$E\{X_{t_1}X_{t_2}\} = \sum_{n=1}^{k} \sigma^2[\cos(n\lambda t_1)\cos(n\lambda t_2) + \sin(n\lambda t_1)\sin(n\lambda t_2)].$$

From our previous trigonometric identities,

$$E\{X_{t_1}X_{t_2}\} \;=\; \sum_{n=1}^{k}\sigma^2\cos\Big(n\,\lambda\,(t_2-t_1)\Big).$$

This process is thus weakly stationary. Also since a linear combination of normal random variables is normal, the process is also normal and hence it is also strictly stationary.

In the discussion which follows, we shall assume that the reader is familiar with Riemann-Stieltjes integration. There is a discussion of Riemann-Stieltjes integration in Appendix A.

A bivariate function $\gamma(s, t)$ is *non-negative definite* if for any set of time points $t_1,\ldots,t_n$ and any set of complex numbers $z_1, \ldots, z_n$

(6.1.3) $$\sum_{j=1}^{n}\;\sum_{k=1}^{n}\gamma\,(t_j,\;t_k)\,z_j\,z_k^{*}\;\geq\;0.$$

If the inequality is strict, the function is said to be *positive definite*. In general we will use the * to indicate the complex conjugate. That is if z is a complex number given by $a + bi$, the $z^*$ is $a - bi$. Notice that if $\gamma(s, t) = E(X_sX_t^*)$ is the covariance function of a zero mean process, then

(6.1.4) $$\sum_{j=1}^{n}\sum_{k=1}^{n}\gamma(t_j,\;t_k)\,z_j\,z_k^{*}\;=\;E\Big\{\sum_{j=1}^{n}\;\sum_{k=1}^{n}X_{t_j}\;X_{t_k}^{*}z_jz_k^{*}\Big\}\;=\;E\left[\sum_{j=1}^{n}X_{t_j}z_j\sum_{k=1}^{n}X_{t_k}^{*}z_k^{*}\right].$$

Rewriting (6.1.4) as a modulus square, we have

(6.1.5) $$\sum_{j=1}^{n}\sum_{k=1}^{n}\gamma(t_j,\;t_k)\,z_j\,z_k^{*}=\;E\mid\sum_{j=1}^{n}X_{t_j}z_j\mid^2\;\geq\;0.$$

Thus the covariance function of a zero-mean stochastic process is a non-negative definite function. A trivial modification shows, in fact, that every covariance function is non-negative definite. The converse is also true, i.e. if $\gamma$ is non-negative definite, then there is a stochastic process for which it is a covariance function.

If we assume $X_t$ is complex valued and stationary, then without any loss of generality, we may assume $E(X_t) = 0$. Notice in this case as we did in Chapter 1, we have

(6.1.6) $$\gamma\,(s, t)\;=\;E\,(X_s\,X_t^{*})\;=\;E\,(X_{s+\tau}\,X_{t+\tau}^{*}).$$

Now let $\tau = -t$ so that $\gamma(s, t) = E(X_{s-t}\,X_0^{*}) = \gamma(s - t, 0)$. It is customary to write $\gamma_t = \gamma(t, 0)$ Thus we may define a one-dimensional function $\gamma_t$ to be non-negative definite if and only if $\gamma(t, t') = \gamma_{t-t'}$ is non-negative definite. Notice that if $\gamma_s = E(X_{t+s}\,X_t^{*})$, then

(6.1.7) $$\gamma_{-s}\;=\;E(X_{t-s}\,X_t^{*})\;=\;E(X_t\,X_{t+s}^{*})\;=\;\gamma_s^{*}.$$

The property $\gamma_{-s} = \gamma_s^{*}$ we have encountered before in connection with $J(-\lambda) = \overline{J(\lambda)}$. This is called the *Hermetian property*. If the times series is real-valued, then, of course, the Hermetian property becomes *a symmetry property.*

Because $\gamma_{-s} = \gamma_s$, the covariance matrix associated with this stationary, real-valued time series is,

$$\begin{bmatrix}\gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} & \gamma_{n-1}\\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3} & \gamma_{n-2}\\ \gamma_2 & \gamma_1 & \cdots & \gamma_{n-4} & \gamma_{n-3}\\ \vdots & & & & \vdots\\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_1 & \gamma_0\end{bmatrix}\;=\;\Gamma_n.$$

Note that the matrix is symmetric with constant elements on any diagonal. The corresponding correlation matrix is $R_n = \Gamma_n/\gamma_0$.

One of the consequences of positive definiteness is that the determinant and all of its principal minors are greater than 0.

Hence

$$\begin{vmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{vmatrix} > 0$$

or

$$\gamma_0^2 - \gamma_1^2 > 0$$

or

$$\gamma_0 > |\gamma_1|$$

or $|\rho_1| < 1$ where $\rho_s$ is the correlation of lag s.

Similarly,

$$\begin{vmatrix} \gamma_0 & \gamma_1 & \gamma_2 \\ \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_2 & \gamma_1 & \gamma_0 \end{vmatrix} > 0; \quad \begin{vmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{vmatrix} > 0; \text{ and } \begin{vmatrix} \gamma_0 & \gamma_2 \\ \gamma_2 & \gamma_0 \end{vmatrix} > 0.$$

which imply respectively

$$-1 < \frac{\gamma_0\gamma_2 - \gamma_1^2}{\gamma_0^2 - \gamma_1^2} < 1; \quad |\gamma_1| < \gamma_0; \quad |\gamma_2| < \gamma_0$$

or equivalently

$$-1 < \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} < 1; \quad |\rho_2| < 1; \quad |\rho_2| < 1.$$

### 6.2 Basic Spectral Theorems

The discussion we have had earlier focussed on the the discrete Fourier transform of the raw data. Because the raw data itself is a stochastic process, its Fourier transform will also be a stochastic process. While this is interesting from a probabilistic point of view, the fact that our putative spectral distribution is a random process makes it somewhat unsatisfactory for being used a diagnostic tool. In order to stabilize the notion of a spectrum, we need to take the Fourier transform of a deterministic function. Thus, we would like to replace $X_t$ with the covariance function, $\gamma_\tau$. The following series of results are designed to lead us to understanding and interpreting the transform of the covariance function.

**Bochner's Theorem:** A continuous function $\gamma(t)$ is non-negative definite if and only if it can be represented by an integral of the form

$$\gamma(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\, d\, F(\lambda) \tag{6.2.1}$$

where G is a real-valued, non-decreasing and bounded function. The function, F, will be the analog of a probability distribution function as we shall see shortly. Notice that (6.2.1) strongly resembles relationship for the inverse Fourier transform we had discussed earlier. Recall

$$X_t = \frac{n}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda t}\, J(\lambda)\, d\lambda. \tag{6.2.2}$$

The proof of Bochner's Theorem is beyond the scope of these lecture notes and will be left as an unproven assertion. Clearly to employ Bochner's Theorem as it might apply to a covariance function, $\gamma_\tau$, we need only show that $\gamma_\tau$ is continuous.

**Lemma:** If $\gamma_\tau$ is continuous at $\tau = 0$, then it is continuous at all $\tau$.
Proof: We first consider $\gamma$ evaluated at $\tau$ and $\tau'$. Now

$$|\gamma_\tau - \gamma_{\tau'}| = |E\left\{ X_{t+\tau} X_t^* \right\} - E\left\{ X_{t+\tau'} X_t^* \right\}| = | E\left\{ (X_{t+\tau} - X_{t+\tau'}) X_t^* \right\}|$$

Since $X_t$ is stationary, we have by the Schwartz Inequality,

$$|\gamma_\tau - \gamma_{\tau'}| = |E\left\{ (X_\tau - X_{\tau'}) X_0^* \right\}| \le \left( E\left\{ | X_0^* |^2 \right\} \right)^{\frac{1}{2}} \left( E\left\{ | X_\tau - X_{\tau'}|^2 \right\} \right)^{\frac{1}{2}}.$$

But $E\left\{ |X_\tau - X_{\tau'}|^2 \right\} = E\left\{ X_\tau X_\tau^* \right\} - E\left\{ X_{\tau'} X_\tau^* \right\} - E\left\{ X_\tau X_{\tau'}^* \right\} + E\left\{ X_{\tau'} X_{\tau'}^* \right\} = 2\gamma_0 - \gamma_{\tau-\tau'} - \gamma_{\tau'-\tau}$.
As $\tau - \tau' \to 0$, $\gamma_{\tau-\tau'} \to \gamma_0$ and $\gamma_{\tau'-\tau} \to \gamma_0$ by assumption. Hence $E\left\{ | X_\tau - X_{\tau'}|^2 \right\} \to 0$ so that $|\gamma_\tau - \gamma_{\tau'}| \to 0$. Thus $\gamma_\tau$ is continuous for all $\tau$.

We are now in a position to combine Bochner's Theorem with the basic Lemma we have just show to obtain the basic spectral representation theorem.

**Spectral Representation Theorem:** If $\gamma_\tau$ is the covariance function of $X_t$ and $\gamma_\tau$ is continuous at 0, then there is a real-valued, non-decreasing bounded function F such that

$$\gamma_\tau = \int_{-\infty}^{\infty} e^{i\lambda\tau}\, d\, F(\lambda) \tag{6.2.3}$$

The function, F, is called the *spectral distribution function.* The proof follows from Bochner's Theorem and the fact that $\gamma_\tau$ is non-negative definite and continuous.
Notice that $\gamma_0 = \int_{-\infty}^{\infty} d\, F(\lambda) = F(\infty) - F(-\infty) = \text{var}\,(X_t)$. Since F is defined up to an additive constant, choose $F(-\infty) = 0$, so that $\gamma_0 = F(\infty)$. As a convention, we choose $F(\lambda) = F(\lambda + )$, i.e. we assume continuity on the right. If F has a derivative, then $f(\lambda) = F'(\lambda) = \frac{d\, F(\lambda)}{d\, \lambda}$ and

$$\gamma_\tau = \int_{-\infty}^{\infty} e^{\,i\lambda\tau}\, f(\lambda)\, d\lambda. \tag{6.2.4}$$

The function $f(\lambda)$ is called the *spectral density* of $X_t$. The set of all discontinuities of F is called the *point spectrum* of $X_t$.

The spectral representation theorem we have just seen express the covariance function in terms of the spectral distribution function or the spectral density function. Of major interest is the situation when the spectral distribution or the spectral density can be represented in terms of the covariance function.

**Inverse Fourier Transform Theorem:** If $\int_{-\infty}^{\infty} |\gamma_\tau| \, d\tau < \infty$, then

$$f(\lambda) = F'(\lambda) \text{ exists and } f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda t} \gamma_t \, d\, t. \tag{6.2.5}$$

Again we shall leave this inverse result unproven. Recall, however,

$$J(\lambda) = \frac{1}{n} \sum_{j=0}^{n-1} e^{-i\lambda t} X_t. \tag{6.2.6}$$

This is a rather striking analogy. Of course, we are dealing with continuous-time processes in the present discussion. If $X_t$ is not a continuous-time process but rather a discrete-time process, the analogies are even more striking.

To see this, let $\gamma_j$ be the covariance function for a wide sense stationary discrete time process, $X_t$, $t = 0, \pm 1, \pm 2, \ldots$. The discrete analog of the Spectral Representation Theorem is known as the Herglotz Lemma and is given below.

**Herglotz Lemma:** If $\gamma_j$ is the covariance function indicated above, then there is a real-valued, non-decreasing, bounded function, F, defined on the interval $(-\pi, \pi)$ such that

$$\gamma_j = \int_{-\pi}^{\pi} e^{i\lambda j} d\, F(\lambda). \tag{6.2.7}$$

Recall in the case of the discrete Fourier transform we studied earlier that

$$X_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda t} \, n \, J(\lambda) \, d\lambda \tag{6.2.8}$$

so that we have an even more striking analogy. As in the continuous case, F is defined up to an additive constant. Choose $F(-\pi) = 0$, so that $F(\pi) = \gamma_0$ and as before we take $F(\lambda) = F(\lambda + )$, i.e. choose F to be continuous from the right. The discrete version of the inverse Fourier transform can then be given.

**Inverse Fourier Transform Theorem (Discrete Version):** If $\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty$, then

$$f(\lambda) = F'(\lambda) \text{ exists and } f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} e^{-i\lambda j} \gamma_j, \quad |\lambda| < \pi. \tag{6.2.9}$$

Recall for the discrete Fourier transform we had

$$J(\lambda) = \frac{1}{n} \sum_{j=0}^{n-1} e^{-i\lambda j} X_j. \tag{6.2.10}$$

To this point, we have been assuming that $X_t$ is a complex-valued process. If, in fact, $X_t$ is a real-valued process, $\gamma_\tau = \gamma_{-\tau}$. In this case

$$\int_{-\infty}^{\infty} e^{i\lambda\tau} \, d\, F(\lambda) = \gamma_\tau = \gamma_{-\tau} = \int_{-\infty}^{\infty} e^{-i\lambda\tau} \, dF(\lambda) \tag{6.2.11}$$

Summing the expressions for $\gamma_\tau$ and $\gamma_{-\tau}$, we have

$$2\gamma_\tau \;=\; \gamma_\tau + \gamma_{-\tau} \;=\; \int_{-\infty}^{\infty} (e^{\,i\lambda\tau} + e^{-i\lambda\tau})\, d\, F(\lambda)\,.$$

Thus

$$\gamma_\tau = \int_{-\infty}^{\infty} \frac{(e^{\,i\lambda\tau}+e^{-i\lambda\tau})}{2}\, d\, F(\lambda) \;=\; \int_{-\infty}^{\infty} \cos(\lambda\tau)\, d\, F(\lambda). \tag{6.2.12}$$

This implies that $\int_{-\infty}^{\infty} \sin(\lambda\tau)\, d\, F(\lambda) \;=\; 0$. Similarly

$$f(\lambda) \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda t}\gamma_t\, d\, t \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} \cos(-\lambda t)\, \gamma_t\, d\, t. \tag{6.2.13}$$

Because of the even symmetry of the cosine function

$$f(\lambda) \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} \cos(\lambda t)\, \gamma_t\, d\, t \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\lambda t}\gamma_t\, d\, t \;=\; f(-\lambda). \tag{6.2.14}$$

Thus for a real-valued stochastic process, the spectral density is symmetric about the 0 frequency. A practical implication of this fact is that we really need not plot the spectral density function over the whole range from $-\pi$ to $\pi$. The range from 0 to $\pi$ is sufficient. If $\gamma_\tau$ is real so that $\gamma_\tau$ is symmetric about 0, then we may define $G(\lambda) = 2F(\lambda) - \gamma_0$. In this case

$$\gamma_\tau \;=\; \int_{-\infty}^{\infty} e^{\,i\lambda\tau}\, d\, F(\lambda) \;=\; \int_{-\infty}^{\infty} \cos(\lambda\tau)\, d\, F(\lambda) \;=\; \int_{0}^{\infty} \cos(\lambda\tau)\, d\, G(\lambda). \tag{6.2.15}$$

The function G is sometimes called the *spectral distribution in real form.*

Note that the spectral representation theorem could be (and often is) formulated in many different ways depending on whether or not the constant $2\pi$ is included in the equation representing $\gamma_\tau$ or the equation representing $f(\lambda)$. A few of the more common representations are

$$\gamma_j \;=\; \int_{-\pi}^{\pi} e^{\,i\lambda j}\, d\, F(\lambda) \tag{6.2.16}$$

$$\gamma_j \;=\; \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{\,i\lambda j}\, d\, F(\lambda) \tag{6.2.17}$$

or

$$\gamma_j \;=\; \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i f j}\, d\, F(f). \tag{6.2.18}$$

In this latter expression, $f$ is the so-called natural frequency whereas $\lambda$ is often called the angular frequency. The relationship between the two is $\lambda \;=\; 2\pi f$. All three equations (6.2.16), (6.2.17) and (6.2.18) are equivalent, but not identical. Care must be taken to insure the use of the correct constants.

### 6.3 Interpreting Spectral Densities

We have shown that we may write the covariance function as a Fourier transform of the spectral distribution function (the spectral density) and as we did with the DFT applied to the raw data, we can interpret the spectral density function as indicating the intensity of the sinusoidal components in the covariance function.

That is to say, $f(\lambda)$ is the intensity of the fundamental sinusoid at frequency $\lambda$ in the covariance, $\gamma_\tau$, much the same way that $|J(\lambda)|^2 = R^2(\lambda)$ was the intensity of the fundamental sinusoid at frequency $\lambda$ in the data $X_t$. Of course $J(\lambda)$ is a random process while $f(\lambda)$ is a deterministic function. Of course, as we indicated earlier, this is the reason for exploiting Bochner's Theorem.

We now give a fundamental theorem which will illustrate the relationship between the Fourier transform of $X_t$ and the Fourier transform of $\gamma_\tau$. Clearly we want to have the Fourier transform of $\gamma_\tau$ represent in some sense the structure of the Fourier transform of $X_t$. We will only give the statement of the following theorem.

Let $X(t) = X_t$ be a zero mean (continuous) process with spectral distribution function, $F(\lambda)$. A process $K(\lambda)$ is said to have orthogonal increments if for every set, $\lambda_1 < \lambda_2 \leq \lambda_3 < \lambda_4$, we have

$$\text{(6.3.1)} \qquad E\left\{ \Big(K(\lambda_4) - K(\lambda_3)\Big)\Big(K(\lambda_2) - K(\lambda_1)\Big) \right\} = 0.$$

**Spectral Representation Theorem** For every stationary process, $X(t)$, there is a stochastic process, $K(\lambda)$, with orthogonal increments, such that for each fixed $t$

$$\text{(6.3.2)} \qquad X(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\, d\,K(\lambda),$$

$K(\lambda)$ is defined up to an additive random variable. We may fix this additive random variable by taking $K(-\infty) = 0$. In this case,

$$E\Big( K(\lambda)\Big) = 0,$$
$$E\Big\{ |K(\lambda)|^2 \Big\} = F(\lambda),$$

and

$$E\Big\{ |d\,K(\lambda)|^2 \Big\} = d\,F(\lambda).$$

The process, $K(\lambda)$, is called the *spectral representation process.* (The DFT process, $J(\lambda)$, is an approximation to $K(\lambda)$ based on a time series). This result may be interpreted to say that the expected sinusoidal oscillations in $X_t$ are the same sinusoidal oscillations in $\gamma_\tau$, i.e. the structure we expect to find in $X_t$ are found in $\gamma_\tau$. We say this in the sense that the respective spectral densities are the same. This gives us confidence that using $\gamma_\tau$ as a deterministic representative of $X_t$ is very satisfactory since (at least as far as frequency content is concerned) the structure of $\gamma_\tau$ is reflective of the structure we expect to find for $X_t$.

# Chapter 7
# Estimating Spectral Densities

**7.1 Estimating of the Mean of a Stationary Time Series.** I want to make a slight change of notation here. Instead of writing the time series as $X_0,\ldots,X_{n-1}$, I will write the series as $X_1,\ldots,X_n$. Previously, in our discussion of complex sinusoids for, has been convenient to start with an index of zero so as to make the first term in the DFT series one. For reasons of symmetry in our present discussion it will be more convenient to make the first index one. The major effect of this shift is to create a phase shift in the computation of the DFT, i.e. multiplication by a complex constant whose modulus is one. This will be irrelevant to the use of the DFT as a diagnostic tool.

As before, let $X_t$ be a real-valued stationary process with constant mean, $\mu$, and covariance function, $\gamma_\tau$. We assume that we have real-valued time series $X_1,\ldots,X_n$. We consider the sample mean, $\overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$. Notice that $E(\overline{X}) = \frac{1}{n}E\left\{\sum_{t=1}^{n} X_t\right\} = \frac{1}{n}\sum_{t=1}^{n}\mu = \mu$ so that $\overline{X}$ is an unbiased estimate of $\mu$. Recall that $\overline{X}$ is not the maximum likelihood estimate for the mean, $\mu$, in the hidden periodicities model. Here, however, we are only prepared to assume a stationary model for the process. Thus we are not yet able to compute all other properties of $\overline{X}$. We are interested in computing the variance of $\overline{X}$. Unlike the usual independent and identically distributed case (i.i.d.), the computation is not straightforward. We start with a lemma.

**Lemma:** If $a_j$, $j = 0, \pm 1, \pm 2, \ldots$, then

$$\sum_{t=1}^{m}\sum_{s=1}^{m} a_{t-s} = \sum_{t=-m+1}^{m-1} (m - |t|)\, a_t. \tag{7.1.1}$$

Similarly, if h(t) is a real-valued function,

$$\int_0^T\int_0^T h(t-s)\, ds\, dt = \int_{-T}^{T} (T - |t|)\, h(t)\, dt. \tag{7.1.2}$$

**Proof:** We prove the discrete case. The other follows in a similar way. Consider

$$\sum_{t=1}^{m}\sum_{s=1}^{m} a_{t-s} = \sum_{t=1}^{m}\sum_{j=t-m}^{t-1} a_j$$

where we are letting $t - s = j$. Interchanging the order of summation (see **** figure), we obtain

$$\sum_{t=1}^{m}\sum_{s=1}^{m} a_{t-s} = \sum_{j=-m+1}^{0}\sum_{t=1}^{j+m} a_j + \sum_{j=1}^{m-1}\sum_{t=j+1}^{m} a_j.$$

Rewriting yields

$$\sum_{t=1}^{m}\sum_{s=1}^{m} a_{t-s} = \sum_{j=-m+1}^{0} (m + j)\, a_j + \sum_{j=1}^{m-1} (m - j)\, a_j = \sum_{j=-m+1}^{m-1} (m - |j|)\, a_j.$$

This gives us the result in the discrete case. the continuous case follows in a similar way although actually the computation is slightly easier.

Applying this lemma to the computation of the variance of $\overline{X}$, we have

$$\text{var}(\bar{X}) = E\left\{(\bar{X} - \mu)^2\right\} = E\left\{\frac{1}{n}\sum_{t=1}^{n}(X_t - \mu)\right\}^2 = \frac{1}{n^2} E\left\{\sum_{t=1}^{n}\sum_{s=1}^{n}(X_t - \mu)(X_s - \mu)\right\}.$$

Then

$$\text{var}(\bar{X}) = \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\gamma_{t-s} = \frac{1}{n^2}\sum_{t=-n+1}^{n-1}(n - |t|)\gamma_t.$$

Canceling an $n$ yields,

$$\text{var}(\bar{X}) = \frac{1}{n}\sum_{t=-n+1}^{n-1}(1 - \frac{|t|}{n})\gamma_t. \tag{7.1.3}$$

We are in a position to have the following theorem.

**Theorem:** Let $\bar{X} = \frac{1}{n}\sum_{t=1}^{n}X_t$. Then $\bar{X}$ is unbiased for $\mu$ with variance equal to

$$\frac{1}{n}\sum_{t=-n+1}^{n-1}\left\{1 - \frac{|t|}{n}\right\}\gamma_t.$$

If $\sum_{t=-\infty}^{\infty}|\gamma_t| < \infty$, then

$$\text{var}(\bar{X}) \sim \frac{1}{n}\sum_{t=-\infty}^{\infty}\gamma_t \text{ as } n \to \infty. \tag{7.1.4}$$

In particular, $\text{var}(\bar{X}) \to 0$ as $n \to \infty$. Further if $X_t$ is a discrete process,

$$\sum_{t=-\infty}^{\infty}\gamma_t = 2\pi f(0) \tag{7.1.5}$$

where $f$ is the spectral density of $X_t$. The expression, $\sim$, means "is asymptotic to."

**Proof:** The first part has already been shown. To see the second part, consider

$$n\,\text{var}(\bar{X}) = \sum_{t=-n+1}^{n-1}\left(1 - \frac{|t|}{n}\right)\gamma_t = \sum_{t=-n+1}^{n-1}\gamma_t - \frac{1}{n}\sum_{t=-n+1}^{n-1}|t|\,\gamma_t.$$

The second term can be rewritten as

$$\frac{1}{n}\sum_{t=-n+1}^{n-1}|t|\,\gamma_t = \sum_{t=-\infty}^{\infty}I_n(t)\frac{|t|}{n}\gamma_t$$

where term by term we define $I_n$ by,

$$I_n(t) = \begin{cases} 1 & \text{if } -n+1 \le t \le n-1 \\ 0 & \text{otherwise} \end{cases}.$$

Notice that $|I_n(t)\frac{|t|}{n}\gamma_t| \le |\gamma_t|$. By results on series from advanced calculus, we know

$$\lim_{n\to\infty} \sum_{t=-\infty}^{\infty} I_n(t)\, \tfrac{|t|}{n}\, \gamma_t \;=\; \sum_{t=-\infty}^{\infty} \lim_{n\to\infty} I_n(t)\, \tfrac{|t|}{n}\, \gamma_t \;=\; \sum_{t=-\infty}^{\infty} 0 \;=\; 0.$$

Hence $n\,\mathrm{var}(\bar{X}) \to \sum_{t=-\infty}^{\infty} \gamma_t + 0 = \sum_{t=-\infty}^{\infty} \gamma_t$. Finally note that

$$f(\lambda) \;=\; \tfrac{1}{2\pi} \sum_{t=-\infty}^{\infty} e^{-i\lambda t}\, \gamma_t,$$

so that

$$f(0) \;=\; \tfrac{1}{2\pi} \sum_{t=-\infty}^{\infty} e^{-0}\gamma_t \;=\; \tfrac{1}{2\pi} \sum_{t=-\infty}^{\infty} \gamma_t.$$

Thus $n\,\mathrm{var}(\bar{X}) \to 2\pi f(0)$. Similar results hold for the continuous case. Specifically

(7.1.6) $$E(\bar{X}) = \mu,$$

(7.1.7) $$\mathrm{var}(\bar{X}) \;=\; \tfrac{1}{T^2} \int_{-T}^{T} (1 - \tfrac{|t|}{T})\, \gamma_t \, dt$$

with

(7.1.8) $$T\,\mathrm{var}(\bar{X}) \;\to\; \int_{-\infty}^{\infty} \gamma_t \, dt \;=\; 2\pi f(0).$$

**7.2 Estimation of Variances and Covariances.** Let $X_1,\ldots,X_n$ be a time series from $X_t$ which is stationary process with mean $\mu$ and covariance function $\gamma_\tau$. We want now to estimate $\gamma_\tau$. A first approximation variance estimator might be

(7.2.1) $$\tilde{\gamma}_0 \;=\; \tfrac{1}{n} \sum_{t=1}^{n} (X_t - \bar{X})^2$$

the usual estimator of variance with the general stationary covariance estimator

(7.2.2) $$\tilde{\gamma}_\tau \;=\; \tfrac{1}{n-\tau} \sum_{t=1}^{n-\tau} (X_t - \bar{X})(X_{t+s} - \bar{X}).$$

This, like other ad hoc estimators, has some potential drawbacks. One might object that the first factor in the summation given in (7.2.2) involves the observations $X_1,\ldots,X_{n-s}$ and yet we subtract $\bar{X}$ which is computed from the whole set of observations, $X_1,\ldots,X_n$. Similarly, the second factor involves $X_{s+1},\ldots,X_n$ and again we subtract $\bar{X}$. In spite of the symmetry, this clearly will cause the cross products we expect to cancel in quadratic form computations not to cancel. Hence, a second proposed estimator might be

(7.2.3) $$\bar{\gamma}_s \;=\; \tfrac{1}{n-s} \sum_{t=1}^{n-s} (X_t - \bar{X}_{1,s})(X_{t+s} - \bar{X}_{2,s})$$

where $\bar{X}_{1,s} = \tfrac{1}{n-s} \sum_{t=1}^{n-s} X_t$ and $\bar{X}_{2,s} = \tfrac{1}{n-s} \sum_{t=s+1}^{n} X_t$. In this case, expanding the cross product, we have

(7.2.4) $$E(\bar{\gamma}_s) \;=\; \tfrac{1}{n-s}\, E\left( \sum_{t=1}^{n-s} \left\{ X_t X_{t+s} - \bar{X}_{1,s} X_{t+s} - \bar{X}_{2,s} X_t + \bar{X}_{1,s}\bar{X}_{2,s} \right\} \right)$$

so that

(7.2.5) $$E(\overline{\gamma}_s) = \frac{1}{n-s}E\left( \sum_{t=1}^{n-s} X_t X_{t+s} - 2(n-s)\overline{X}_{1,s}\overline{X}_{2,s} + (n-s)\overline{X}_{1,s}\overline{X}_{2,s} \right).$$

Canceling similar terms, we have

(7.2.6) $$E(\overline{\gamma}_s) = \frac{1}{n-s}E\left( \sum_{t=1}^{n-s} X_t X_{t+s} - (n-s)\overline{X}_{1,s}\overline{X}_{2,s} \right).$$

This may be rewritten as

(7.2.7) $$E(\overline{\gamma}_s) = E\left( \frac{1}{n-s}\sum_{t=1}^{n-s}(X_t - \mu)(X_{t+s} - \mu) + \mu\overline{X}_{1,s} + \mu\overline{X}_{2,s} - \mu^2 - \overline{X}_{1,s}\overline{X}_{2,s} \right)$$

which in turn yields

(7.2.8) $$E(\overline{\gamma}_s) = E\left( \frac{1}{n-s}\sum_{t=1}^{n-s}\left\{ (X_t - \mu)(X_{t+s} - \mu) \right\} - (\overline{X}_{1,s} - \mu)(\overline{X}_{2,s} - \mu) \right).$$

Finally this becomes

(7.2.9) $$E(\overline{\gamma}_s) = \gamma_s - \text{cov}(\overline{X}_{1,s}, \overline{X}_{2,s}).$$

But

(7.2.10) $$\text{cov}(\overline{X}_{1,s}, \overline{X}_{2,s}) = E\left\{ \sum_{j=1}^{n-s} \frac{(X_j - \mu)}{n-s} \sum_{i=1}^{n-s} \frac{(X_{i+s} - \mu)}{n-s} \right\}$$

which in turn may be written

$$\text{cov}(\overline{X}_{1,s}, \overline{X}_{2,s}) = \frac{1}{(n-s)^2}\sum_{j=1}^{n-s}\sum_{i=1}^{n-s} E(X_j - \mu)(X_{i+s} - \mu)$$

so that

$$\text{cov}(\overline{X}_{1,s}, \overline{X}_{2,s}) = \frac{1}{(n-s)^2}\sum_{j=1}^{n-s}\sum_{i=1}^{n-s}\gamma_{s+i-j}.$$

Using the (7.1.1) from the Lemma, we have

(7.2.11) $$\text{cov}(\overline{X}_{1,s}, \overline{X}_{2,s}) = \frac{1}{(n-s)^2}\sum_{j=-n+s+1}^{n-s-1}\gamma_{s+j}(n - s - |j|).$$

As before, if $\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty$, then $\text{cov}(\overline{X}_{1,s}, \overline{X}_{2,s}) \to 0$ as $n \to \infty$. We may compute the asymptotic bias by using (7.2.9) to obtain

(7.2.12) $$n[E(\overline{\gamma}_s) - \gamma_s] = -n\,\text{cov}(\overline{X}_{1,s}, \overline{X}_{2,s}) = \frac{-n}{n-s}\sum_{j=-n+s+1}^{n-s-1}(1 - \frac{|j|}{n-s})\gamma_{s+j}.$$

Letting n $\to \infty$, we have

(7.2.13) $$n[E(\bar{\gamma}_s) - \gamma_s] \to -\sum_{j=-\infty}^{\infty} \gamma_{s+j} = \gamma_s.$$

This shows that $\bar{\gamma}_s$ is asymptotically unbiased. Computing the asymptotic mean of $\bar{\gamma}_s$ is not a trivial task. Computing the corresponding mean of $\tilde{\gamma}_s$ is a real mess! We will not show the algebra of the computation, but for the record,

(7.2.14) $$E(\tilde{\gamma}_s) = \gamma_s + \text{var}(\bar{X})$$

$$- \tfrac{1}{n(n-s)} \left\{ \sum_{k=s+1}^{n-1} (n-k) + \sum_{k=1}^{s} (n-s) + \sum_{k=-(n-s)-1}^{0} (n-s+k) \right\} (\gamma_k + \gamma_{k+s}).$$

Taking limits as n $\to \infty$ and simplifying

(7.2.15) $$E(\tilde{\gamma}_s) \sim \gamma_s + \text{var}(\bar{X}) - \tfrac{1}{n} r_n$$

where $\lim_{n \to \infty} r_n = 2 \sum_{k=-\infty}^{\infty} \gamma_k$, so that

(7.2.16) $$n\left(E(\tilde{\gamma}_s) - \gamma_s\right) \to -\sum_{j=-\infty}^{\infty} \gamma_k.$$

Thus $E(\tilde{\gamma}_s) - \gamma_s \sim \text{var}(\bar{X}) \neq 0$ as $n \to \infty$. In general $\tilde{\gamma}_s$ is not asymptotically unbiased.

A simpler estimator assumes we start with a zero mean process. In this case a natural estimator of $\gamma_s$ is

(7.2.17) $$\hat{\gamma}_s = \tfrac{1}{n} \sum_{t=1}^{n-s} X_t X_{t+s}.$$

It is straightforward to see

(7.2.18) $$E(\hat{\gamma}_s) = \tfrac{1}{n} \sum_{t-1}^{n-s} \gamma_s = \gamma_s \tfrac{n-s}{n} \to \gamma_s \text{ as } n \to \infty.$$

Whereas the computation of the variance of $\tilde{\gamma}_s$ and $\bar{\gamma}_s$ is all but impossible, some results are known about the variance of $\hat{\gamma}_s$. We may get some idea by considering a special case. The following results were developed in ****. Let $X_t = \sigma \sum_{j=0}^{\infty} b_j \eta_{t-j}$. We shall later see this as an *infinite moving average* where $\eta_t$ are independent random variables with $E(\eta_t) = 0$, $\text{var}(\eta_t) = 1$, $E(\eta_t^4) < \infty$ and $|b_j| < Kc^j$, $0 \leq c<1$. Let $\hat{\rho}_s = \frac{\hat{\gamma}_s}{\hat{\gamma}_0}$. Then

(7.2.19) $$\lim_{n \to \infty} n\left\{E(\hat{\rho}_s - \rho_s\right\} = 2 \sum_{j=-\infty}^{\infty} (\rho_j^2 \rho_s - \rho_j \rho_{j-s}).$$

To compute second moments of $X_t$, we have

(7.2.20) $$\lim_{n \to \infty} n \text{ var } (\hat{\rho}_s) = \sum_{j=-\infty}^{\infty} (\rho_j^2 + \rho_{j+s} \rho_{j-s} - 4 \rho_j \rho_s \rho_{j-s} + 2 \rho_s^2 \rho_j^2) \stackrel{d}{=} V_s$$

and

(7.2.21) $$\lim_{n\to\infty} \text{n cov}(\hat{\rho}_s, \hat{\rho}_{s+v}) = \sum_{j=-\infty}^{\infty} (\rho_j\rho_{j+v} + \rho_{j+s+v}\rho_{j-s} - 2\,\rho_{s+v}\rho_j\rho_{j-s} - 2\,\rho_j\rho_s\rho_{j-s-v} + 2\,\rho_s\rho_{s+v}\rho_j^2) = U_{s,v}.$$

From this the correlation is seen to be,

(7.2.22) $$\lim_{n\to\infty} \text{corr}(\hat{\rho}_s, \hat{\rho}_{s+v}) = \frac{U_{s,v}}{V_s}.$$

The asymptotic distribution is calculated by

(7.2.23) $$\sqrt{n}\,(\hat{\rho}_s - \rho) \xrightarrow{\text{dist}} n(0, V_s).$$

Thus $\hat{\rho}_s$ is asymptotically unbiased, consistent and asymptotically normal. Under appropriate assumptions, $\hat{\rho}_s$ (and $\hat{\gamma}_s$ by implication) seem to be reasonable estimators of respectively the correlation, $\rho_s$, and the covariance, $\gamma_s$, respectively. Notice that is we consider $\lim_{s\to\infty} V_s$, then we have

(7.2.24) $$\lim_{s\to\infty} V_s = \lim_{s\to\infty}\lim_{n\to\infty} \text{n var}(\hat{\rho}_s) = \sum_{j=-\infty}^{\infty} \rho_j^2.$$

Hence even if n is large, when we try to estimate $\rho_s$ for large lags, the variance of the estimate $\hat{\rho}_s$ does not converge to 0.

**7.3 The Periodogram and Other Smoothers.** We restrict our attention to real-valued, wide-sense stationary time series with a spectral density function. Let $X_1,\ldots,X_n$ be the time series. We want to estimate the spectral density function, $f(\lambda)$, $|\lambda| \le \pi$. We assume in this discussion that $\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty$. Recall by definition

(7.3.1) $$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j\, e^{-i\lambda j}.$$

By analogy to (7.3.1), we can consider an empirical density estimator called the periodogram and given as

(7.3.2) $$\frac{1}{2\pi} \sum_{j=-n+1}^{n-1} \hat{\gamma}_j\, e^{-i\lambda j},$$

where $\hat{\gamma}_j = \frac{1}{n} \sum_{t=1}^{n-j} X_t X_{t+j}$ is the estimator of the covariance function, $\gamma_j$. From our earlier discussions, we considered a spectral estimator defined by $R^2(\lambda) = J(\lambda)J^*(\lambda) = |\sum_{t=1}^{n} X_t e^{-i\lambda t}|^2$. To see that these are actually closely related results, we have the following theorem.

**Periodogram Theorem:**

(7.3.3) $$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t\, e^{-i\lambda t} \right|^2 = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} \hat{\gamma}_j\, e^{-i\lambda j}, \quad |\lambda| \le \pi \text{ and}$$

(7.3.4) $$\hat{\gamma}_j = \int_{-\pi}^{\pi} e^{i\lambda j}\, I_n(\lambda)\, d\lambda.$$

**Proof:** Consider second part, (7.3.4), first. Assume that $I_n(\lambda) = \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \hat{\gamma}_k e^{-i\lambda k}$ so that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda j} I_n(\lambda)\, d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda j} \sum_{k=-n+1}^{n-1} \hat{\gamma}_k e^{-i\lambda k}\, d\lambda$$

Interchanging order of summation and integration, we have that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda j} I_n(\lambda)\, d\lambda = \sum_{k=-n+1}^{n-1} \hat{\gamma}_k \frac{\int_{-\pi}^{\pi} e^{i\lambda(k-j)} d\lambda}{2\pi} = \sum_{k=-n+1}^{n-1} \hat{\gamma}_k \delta_{kj} = \hat{\gamma}_j.$$

Here $\delta_{kj}$ is the Dirac delta function. That this integral is either 0 or 1 follows from the orthogonality result (3.1.16). Hence, second part follows immediately from the (7.3.2). Now consider,

$$I_n(\lambda) = \frac{1}{2\pi n} |\sum_{t=1}^{n} X_t e^{-i\lambda t}|^2 = \frac{1}{2\pi n} \left\{ \sum_{t=1}^{n} X_t e^{-i\lambda t} \sum_{s=1}^{n} X_s^* e^{+i\lambda s} \right\}.$$

Simplifying, we have

$$I_n(\lambda) = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s=1}^{n} e^{-i\lambda(t-s)} X_t X_s^*.$$

Now letting $s = t - \tau$ so that $\tau = t - s$ we have

$$I_n(\lambda) = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{\tau=t-1}^{t-n} e^{-i\lambda\tau} X_t X_{t-\tau}^*.$$

We want to interchange summations. For $\tau \leq 0$, the range on t is 1 to $\tau + n$. For $\tau > 0$, the range on t is $\tau + 1$ to n so that

$$In(\lambda) = \frac{1}{2\pi} \sum_{\tau=n+1}^{n-1} e^{i\lambda\tau} \left\{ \frac{1}{n} \sum_{t} X_t X_{t-\tau}^* \right\}.$$

Here because the range of summation on t is dependent on $\tau$, we have deliberately left the summation ambiguous. For $\tau \leq 0$, we consider $\frac{1}{n} \sum_{t=1}^{n+\tau} X_t X_{t-\tau}^*$. Since $\tau$ is negative $-\tau$ is positive and $\hat{\gamma}_\tau$ in this case has precisely the definition, $\frac{1}{n} \sum_{t=1}^{n+\tau} X_t X_{t-\tau}^*$. For $\tau > 0$, we consider $\frac{1}{n} \sum_{t=\tau+1}^{n} X_t X_{t-\tau}^*$. Let $s = t - \tau$ so that $t = s + \tau$. Then

$$\frac{1}{n} \sum_{t=\tau+1}^{n} X_t X_{t-\tau}^* = \frac{1}{n} \sum_{s=1}^{n-\tau} X_s^* X_{s+\tau}.$$

For real-valued processes $X_s^* = X_s$, so that the above sum is the definition of $\hat{\gamma}_\tau$. In either case,

$$I_n(\lambda) = \frac{1}{2\pi} \sum_{\tau=-n+1}^{n-1} \hat{\gamma}_\tau e^{i\lambda\tau}.$$

We define the function $I_n(\lambda)$ to be the periodogram. Thus the periodogram, the square modulus of the DFT of the data, is also the Fourier transform of the covariance estimate in the formula $\hat{\gamma}_\tau = \frac{1}{n} \sum_{t=1}^{n-\tau} X_t X_{t+\tau}$. We develop this connection in more detail. Letting $X_t = \sum_{j=1}^{m} \left\{ A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t) \right\}$, we know from the quick and dirty formulae that

(7.3.5) $$A_j = \frac{2}{n}\sum_{t=1}^{n} X_t \cos(\lambda_j t) \text{ and } B_j = \frac{2}{n}\sum_{t=1}^{n} X_t \sin(\lambda_j t).$$

Consider then

(7.3.6) $$A_j^2 + B_j^2 = \frac{4}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} X_t X_s\Big\{\cos(\lambda_j s)\cos(\lambda_j t) + \sin(\lambda_j s)\sin(\lambda_j t)\Big\}$$

which may be simplified to

$$A_j^2 + B_j^2 = \frac{4}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} X_t X_s \cos(\lambda_j(t-s)).$$

We have assumed that $X_t$ is real valued. We may take the real part of $A_j^2 + B_j^2$ to have

$$A_j^2 + B_j^2 = \text{Real}\left\{\frac{4}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} X_t X_s e^{i\lambda_j(t-s)}\right\} = \text{Real}\left\{\frac{4}{n^2}\sum_{t=1}^{n} X_t e^{i\lambda_j t}\sum_{s=1}^{n} X_s e^{-i\lambda_j s}\right\}.$$

This is simplified to

$$A_j^2 + B_j^2 = \text{Real}\left\{\frac{4}{n^2}\Big|\sum_{t=1}^{n} X_t e^{i\lambda_j t}\Big|^2\right\} = \frac{4}{n^2}\Big|\sum_{t} X_t e^{i\lambda_j t}\Big|^2.$$

Using the results of the Periodogram Theorem, we find that

(7.3.7) $$A_j^2 + B_j^2 = \frac{4}{n^2} \times 2\pi n \times \left\{\frac{1}{2\pi n}\Big|\sum_{t=1}^{n} X_t e^{-i\lambda_j t}\Big|^2\right\} = \frac{8\pi}{n} I_n(\lambda_j).$$

Hence, it is only a slight obvious generalization to find that

(7.3.8) $$R^2(\lambda) = A_\lambda^2 + B_\lambda^2 = \frac{8\pi}{n} I_n(\lambda).$$

where

(7.3.9) $$I_n(\lambda) = \frac{1}{2\pi n}\Big|\sum_{t=1}^{n} X_t e^{-i\lambda t}\Big|^2$$

or

(7.3.10) $$In(\lambda) = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} \hat{\gamma}_j e^{-i\lambda j}$$

with $A_\lambda = \frac{2}{n}\sum_{t=1}^{n} X_t \cos(\lambda t)$ and $B_\lambda = \frac{2}{n}\sum_{t=1}^{n} X_t \sin(\lambda t)$, and with $\hat{\gamma}_j = \frac{1}{n}\sum_{t=j}^{n-j} X_t X_{t+j}$ where $X_t$ is real valued. Thus we have a fundamental linkage between the generalized model of hidden periodicities, the DFT, the spectral density and the covariance function. This gives additional intuitive support to the use of $R^2(\lambda)$ as a diagnostic tool for fitting the model of hidden periodicities.

Next let us consider some of the statistical properties of the periodogram. It is not hard to see that

(7.3.11) $$E[I_n(\lambda)] = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} e^{-i\lambda j} E[\hat{\gamma}_j] = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} e^{-i\lambda j}\gamma_j\left\{\frac{n-|j|}{n}\right\}.$$

The latter equality follows from a slight adaptation of (7.2.18). A little algebra yields

(7.3.12) $$E[I_n(\lambda)] = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} e^{-i\lambda j}\left\{1-\frac{|j|}{n}\right\}\gamma_j \rightarrow \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} e^{-i\lambda j}\gamma_j = f(\lambda), \text{ as } n\rightarrow\infty.$$

From this it follows that $I_n(\lambda)$ is asymptotically unbiased. Consider the variance of $I_n(\lambda)$. In general, this is not a very tractable calculation so that we consider the special case in which $X_1,\ldots,X_n$ are independent random variables. Then

(7.3.13) $$var[I_n(\lambda)] = E\left\{\frac{1}{2\pi}\sum_{j=-n+1}^{n-1} e^{-i\lambda j}\left(A_j - \left\{1-\frac{|j|}{n}\right\}\gamma_j\right)\right\}^2.$$

Realistically there is not much hope for a simple solution based on (7.3.10), The formulation (7.3.9), however, yields a more tractable computation, so consider

(7.3.14) $$E[\, I_n^2(\lambda)] = E\left\{\frac{1}{4\pi^2 n^2}\,|\sum_{t=1}^{n} X_t\, e^{-i\lambda t}\,|^4\right\}.$$

Expanding the quadruple sums yields

(7.3.15) $$E[I_n^2(\lambda)] = \frac{1}{4\pi^2 n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n}\sum_{v=1}^{n} E[X_t\, X_s\, X_u\, X_v]\, e^{-i\lambda t}\, e^{+i\lambda s}\, e^{-i\lambda u} e^{+i\lambda v}.$$

Now let the argument of the quadruple sum be $E_{stuv}$. Depending on the values of s, t, u and v, we will get different sums as follows.

If $s = t \neq u = v$, $E_{stuv} = \sigma^4$ and there are $n(n-1)$ terms of this form.
If $s = u \neq t = v$, $E_{stuv} = \sigma^4 e^{2i\lambda(s-t)}$ and there are $n(n-1)$ terms of this form.
If $s = v \neq t = u$, $E_{stuv} = \sigma^4$ and there are $n(n-1)$ terms of this form.
If $s = t = u = v$, $E_{stuv} = \mu_4$ and there are n terms of this form.
Otherwise, $E_{stuv} = 0$.

We may use these 5 cases to form the equation

(7.3.16) $$E\,[I_n^2(\lambda)] = \frac{1}{4\pi^2 n^2}\left(2\sigma^4 n(n-1) + n\mu^4 + \sigma^4\left\{|\sum_{s=1}^{n} e^{2i\lambda s}|^2 - n\right\}\right)$$

The $-\,n$ term in the expression in braces is meant to account for the case $s = t$ which is double counted in the square modulus. Using our earlier results related to the Dirichlet kernel

(7.3.17) $$E[I_n^2(\lambda)] = \frac{\sigma^4}{2\pi^2} - \frac{\sigma^4}{2\pi^2 n} + \frac{\mu 4}{4\pi^2 n} + \frac{\sigma^4}{4\pi^2}\left(\frac{\sin(n\lambda)}{n\sin(\lambda)}\right)^2 - \frac{\sigma^4}{4\pi^2 n}.$$

Collecting terms we obtain

(7.3.18) $$E[I_n^2(\lambda)] = \frac{\sigma^4}{2\pi^2} + \frac{1}{4\pi^2 n}\left\{\mu_4 - 3\sigma^4\right\} + \frac{\sigma^4}{4\pi^2}\left(\frac{\sin(n\lambda)}{n\sin(\lambda)}\right)^2.$$

But $E[I_n(\lambda)] \rightarrow f(\lambda) = \frac{\sigma^2}{2\pi}$ for a times series which is an independent and identically distributed sequence (called *white noise*). We may combine this with (7.3.18) to obtain an approximation for the variance.

(7.3.19) $$var[I_n(\lambda)] \doteq \frac{\sigma^4}{2\pi^2} + \frac{1}{4\pi^2 n}\left\{\mu_4 - 3\sigma^4\right\} + \frac{\sigma^4}{4\pi^2}\left(\frac{\sin(n\lambda)}{n\sin(\lambda)}\right) - \frac{\sigma^4}{4\pi^2}$$

Simplifying (7.3.19) and recognizing that the second term of (7.3.19) approximates 0 for large n, we obtain

$$\text{var}[I_n(\lambda)] \doteq \frac{\sigma^4}{4\pi^2} + \frac{\sigma^4}{4\pi^2}\left(\frac{\sin(n\lambda)}{n\sin(\lambda)}\right). \tag{7.3.20}$$

If $\lambda = 0$ or $\lambda = \pm\pi$, then $\frac{\sin(n\lambda)}{n\sin(\lambda)} = 1$ and $\text{var}[I_n(\lambda)] \doteq \frac{\sigma^4}{2\pi^2}$ for n large. If $\lambda \neq 0$ and $\lambda \neq \pm\pi$, then $\frac{\sin(n\lambda)}{n\sin(\lambda)} \to 0$ as $n \to \infty$ and $\text{var}[I_n(\lambda)] \doteq \frac{\sigma^4}{4\pi^2}$ for n large.

Thus

$$\text{var}[I_n(\lambda)] \to \begin{cases} \frac{\sigma^4}{4\pi^2}, & \lambda \neq 0 \text{ and } \lambda \neq \pm\pi \\ \\ \frac{\sigma^4}{2\pi}, & \lambda = 0 \text{ or } \lambda \neq \pm\pi. \end{cases} \tag{7.3.21}$$

Thus, even in this simple case of white noise, the variance of the periodogram is asymptotically not equal to zero. In general,

$$\text{var}[I_n(\lambda)] \to \begin{cases} f^2(\lambda), & \lambda \neq 0 \text{ and } \lambda \neq \pm\pi \\ \\ 2\,f^2(\lambda), & \lambda = 0 \text{ or } \lambda = \pm\pi. \end{cases} \tag{7.3.22}$$

Repeating arguments along these lines, we may obtain

$$\text{cov}[I_n(\lambda), I_n(\alpha)] = \left\{ \frac{\sin^2\left(\frac{n}{2}(\lambda-\alpha)\right)}{\sin^2\left(\frac{\lambda-\alpha}{2}\right)} + \frac{\sin^2\left(\frac{n}{2}(\lambda+\alpha)\right)}{\sin^2\left(\frac{\lambda+\alpha}{2}\right)} \right\} \frac{f(\lambda)f(\alpha)}{n^2} \to 0 \text{ as } n \to \infty. \tag{7.3.23}$$

In general, the periodogram is too rough, i.e. its variance is too large. We want to smooth the periodogram. We do this with a kernel smoother.

**7.4 Kernel Smoothers.** We construct a smoothed periodogram by a convolution of the raw periodogram with a smoother function which we will label $K(\lambda)$. Let $K(\lambda)$ be a real-valued, non-negative function such that

$$K(\lambda) = K(-\lambda) \text{ and } \int_{-\infty}^{\infty} K(\lambda)\, d\lambda = 1. \tag{7.4.1}$$

Consider

$$\int_{-\infty}^{\infty} K(\lambda-\alpha)\, I_n(\alpha)\, d\alpha = \int_{-\infty}^{\infty} K(\lambda-\alpha)\, \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} e^{-i\alpha j} \gamma\!\!\!^{\wedge}_j\, d\alpha. \tag{7.4.2}$$

Let $s = \lambda - \alpha$, so that $ds = -d\alpha$ and $\alpha = \lambda - s$.

$$\int_{-\infty}^{\infty} K(\lambda-\alpha)\, I_n(\alpha)\, d\alpha = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} \left\{ \int_{-\infty}^{\infty} K(s)\, e^{isj}\, ds \right\} \gamma\!\!\!^{\wedge}_j e^{-i\lambda j}.$$

If we define $a_j = \int_{-\infty}^{\infty} K(s)\, e^{isj}\, ds$, then

$$\int_{-\infty}^{\infty} K(\lambda-\alpha)\, I_n(\alpha)\, d\alpha = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} a_j \gamma\!\!\!^{\wedge}_j\, e^{-i\lambda j}. \tag{7.4.3}$$

Thus we can weight the covariance function $\gamma_j$ by $a_j$ to smooth $I_n(\lambda)$. We consider therefore estimates of the form

$$\hat{f}_n(\lambda) = \frac{1}{2\pi} \sum_{j=-m}^{m} k(j/m)\, \hat{\gamma}_j\, e^{-i\lambda j} \tag{7.4.4}$$

where $a_j = k(j/m)$ $j = -m, -m+1, \ldots, m-1, m$ and $a_j = 0$ elsewhere. The function, k, is called the *lag window* and m is called the *truncation point.*

$$K_m(\lambda) = \frac{1}{2\pi} \sum_{j=-m}^{m} e^{i\lambda j} k(j/m) \tag{7.4.5}$$

is called the *spectral window.* $K(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\lambda t}\, k(t)\, dt$ is called the *spectral window generator.* Note that $K_m(\lambda) \doteq m\, K(m\lambda)$ if K is concentrated on $(-\pi, \pi)$. We may begin with k(t) and calculate K or $K_m$ as desired. In general, however, we will want to weight the covariance and then use the formula

$$\hat{f}_n(\lambda) = \frac{1}{2\pi} \sum_{j=-m}^{m} k(j/m)\, \hat{\gamma}_j \cos(\lambda j) \tag{7.4.6}$$

as the computational form. We are assuming that the time series $X_t$ is real valued so that the imaginary part of (7.4.4) is zero yielding the very convenient form (7.4.6).

***Examples of lag windows:*** There are several classical windows for smoothing the periodogram. We may think of the smoothed periodogram as a nonparametric spectral density estimator. It is nonparametric because we are not assuming any parametric form for the spectral density. We will later study time-domain models which are indeed parametric. Thus we will also be able to estimate spectral densities parametrically. The smoothed periodogram is in a sense an ad hoc estimator. It is not chosen against any optimality criterion. The practical consideration involves reducing the variance so as to make $f_n(\lambda)$ a statistically consistent estimator of the density $f(\lambda)$. The trade-off is that the bias of $f_n(\lambda)$ will increase. For this reason, a fairly large number of different functional forms for the kernel, k, have been developed.

**Bartlett window.**

$$k(x) = \begin{cases} 1 - |x|, & \text{for } |x| \le 1 \\ 0, & \text{otherwise.} \end{cases} \tag{7.4.7}$$

The Bartlett window is desirable because of its particularly simple computational form.

**Daniell window.**

$$k(x) = \frac{\sin(\pi x)}{\pi x}, \quad -\infty < x < \infty. \tag{7.4.8}$$

**Tukey window.**

$$k(x) = \begin{cases} 1 - 2a + 2a\cos(\pi x), & |x| \le 1 \\ 0, & \text{otherwise.} \end{cases} \tag{7.4.9}$$

For $a = \frac{1}{4}$ and $|x| \le 1$, $k(x) = \frac{1}{2}\left\{1+\cos(\pi x)\right\}$ which, of course, is a particularly simple form. The value $a = .23$ was at one time a recommendation made by John Tukey.

**Parzen window.**

$$k(x) = \begin{cases} 1 - 6x^2 - 6|x|^3, & 0 \le |x| \le .5 \\ 2(1 - |x|)^3, & .5 \le |x| \le 1 \\ 0, & \text{otherwise.} \end{cases} \tag{7.4.10}$$

Generally speaking the choice of a window will have a comparatively minor effect on the spectral density estimator. The choice of the truncation point, m, is more critical An optimal window can actually be computed by minimizing

$$\frac{1}{n}\sum_{j=-n+1}^{n-1}\left\{I_n\left(\frac{j\pi}{n}\right) - f\left(\frac{j\pi}{n}\right)\right\}^2 + \frac{1}{\lambda^{2\nu}\pi}\int_{-\pi}^{\pi}\left\{L\, f(\lambda)\right\}^2 d\lambda \tag{7.4.11}$$

where f is chosen from a class of smooth functions which we shall not attempt to define here The parameter, $\lambda$, is a bandwidth parameter and L is a differential operator. If L is chosen as $\frac{d^2}{d\lambda^2}$, the the solution to (7.4.11) is a cubic spline. The expression is, needless to say a rather complex expression and will not be treated here. A fuller discussion may be found in Wegman (1980, 1981).

**7.5 Statistical Properties of Smoothed Periodograms.** Choice of kernels and truncation points are, of course, factors which will affect the statistical properties of our proposed kernel smoothed periodograms. We state some results here which can be used to establish choices of both.

**Bias.** If both the spectral density, $f(\lambda)$, and and the lag window, $k(\lambda)$, are twice differentiable, then

$$\lim_{m,n\to\infty} m^2\left\{E[f_n(\lambda) - f(\lambda)]\right\} = -k_2 f''(\lambda) \tag{7.5.1}$$

where $k_2 = -\frac{1}{2}k''(0) = \frac{1}{2}\int_{-\infty}^{\infty}\lambda^2 K(\lambda)d\lambda$. For the Tukey window $k_2 = -\frac{\pi^2}{4} \doteq 2.5$. For the Parzen window $k_2 = 6$, so that the Tukey window has better bias properties. Equation (7.5.1) may be rewritten as $E[f_n(\lambda)] - f(\lambda) \doteq -\frac{k_2}{m^2} f''(\lambda)$.

**Variance.** The corresponding expression for the variance can be computed as

$$\lim_{n,m\to\infty} \frac{n}{m}\,\mathrm{var}[f_n(\lambda)] = \begin{cases} f^2(\lambda)\int_{-\infty}^{\infty} k^2(u)\,du, & 0 < |\lambda| < \pi \\ 2\, f^2(\lambda)\int_{-\infty}^{\infty} k^2(u)\,du, & \lambda = 0 \text{ or } \pm\pi. \end{cases} \tag{7.5.2}$$

Thus

$$\mathrm{var}[f_n(\lambda)] \doteq \frac{m}{n} f^2(\lambda)\int_{-\infty}^{\infty} k^2(u)\,du,\ 0 < |\lambda| < \pi. \tag{7.5.3}$$

We want to think of m as a function of n. We write $m_n$. In order to make the smoothed periodogram asymptotically unbiased, using (7.5.1) it is clear that we must choose $m_n \to \infty$. In order that the variance of $f_n(\lambda)$ converge to 0 it is also clear from (7.5.3) that we must choose $\frac{m_n}{n} \to 0$. Moreover, in order that neither the bias nor the variance dominate the mean square error, both the bias squared and the variance must converge to 0 at the same rate. If this happens, then the mean square error will be

$$\mathrm{MSE} = \frac{c_1}{m_n^4} + \frac{c_2 m_n}{n} \tag{7.5.4}$$

where the $c_1$-term represents the bias$^2$ and the $c_2$-term represents the variance. We want these terms to converge to 0 at the same rate (be of the same order) so that we may equate them to obtain

(7.5.5) $$\frac{c_1}{m_n^4} = \frac{c_2 m_n}{n}.$$

In this case $\frac{m_n^5}{n}$ must be a constant so that $m_n^5 = c_3 n$. This implies that $m_n = c_4 n^{\frac{1}{5}}$ is optimal choice of $m_n$ in order to have best rate for mean square error.

**7.6 Practical Spectral Estimation.** There are 3 choices associated with the estimation of the spectral density by a smoothed periodogram, i.e. by

(7.6.1) $$f_n(\lambda) = \frac{1}{2\pi} \sum_{j=-m}^{m} k\left(\frac{j}{m}\right) \hat{\gamma}_j \cos(j\lambda),$$

namely the choice of a lag window, k, the choice of a truncation point, m, and the choice of a set of frequencies, $\lambda_j$ at which to evaluate (7.6.1). We comment on each in turn. For a choice of a lag window, the Tukey window or the Parzen window are usually quite satisfactory. Window choice is not so critical. For the choice of bandwidth, m, there is a much more difficult choice here as m determines the degree of smoothing. The choice of $m_n = c_1 n^{.2}$ is an optimal asymptotic choice, but we usually have a fixed sample size which means that $c_1$ is the critical choice. I recommend that you try $m_n = 10\%$ of n to start, but be prepared for some trial and error. Advanced techniques of cross-validation and other types of error criterion such as Akaike Information Criterion (AIC) can help but are really beyond our discussion here.

The choice of frequencies, $\lambda_j$, at which to evaluate is another fairly critical choice. Here we want a choice that is reasonably dense along the frequency axis. Fourier frequencies $\lambda_j = \frac{2\pi j}{n}$ are not a bad starting point, but are probably too dense for large data sets. Some refinement of that choice with a less fine grid may be in order! In general, there is no optimal theory. This is a nonparametric technique and so some experimentation is in order.

# Chapter 8
# Filtering

**8.1 Linear Filters.** We now turn our attention to filters. A filter maps one stochastic process into another. It can conveniently be and usually is thought of as a black box which has an input and an output respectively, $X_t = X(t)$ as input and $Y_t = Y(t)$ as output. We can use the conceptual black box to model a wide variety of systems such as an amplifier, a communication channel, the ocean or atmosphere or any number of models which may transform one signal into another. Indeed, estimating the filter structure is known in the electrical engineering literature as the system identification problem. From a more mathematical perspective, let $\mathcal{F}$ be a set of complex-valued functions. Let $\phi$: $\mathcal{F} \to \mathcal{F}$. $\phi$ is a *functional* or an *operator* in an abstract mathematical setting. For us $\phi(X_t) = Y_t$ is a filter.

The filter, $\phi$, is *linear* if for every $a_1$, $a_2 \in$ complex numbers, with $X_t^1$, $X_t^2 \in \mathcal{F}$ such that $a_1 X_t^1 + a_2 X_t^2 \in \mathcal{F}$, then $\phi(a_1 X_t^1 + a_2 X_t^2) = a_1\phi(X_t^1) + a_2\phi(X_t^2)$. The filter, $\phi$, is called *time invariant* if for any $X_t \in \mathcal{F}$, then if $X^h(t) = X(t+h)$ and $Y^h(t) = Y(t+h)$ with $\phi(X) = Y$, we have $\phi(X^h) = Y^h$.

**Examples of Filters.** Let $X_t$ be a stochastic process.
**Backward Shift Filter:** $Y_t = X_{t-1}$, $t \in \mathcal{T}$.
**Difference Filter:** $Y_t = X_t - X_{t-1} = \Delta X_t$, $t \in \mathcal{T}$.
**Moving Average Filter:** $Y_t = \sum_{i=0}^{k} a_i X_{t-i}$, $t \in \mathcal{T}$.

**Exponential Filter:** $Y_t = \int_{-\infty}^{t} e^{-\beta(t-s)} X_s \, ds$, $t \in \mathcal{T}$.
**Differential Filter:** $\phi(X_t) = X_t' = \frac{dX_t}{dt}$, $t \in \mathcal{T}$.

**General Linear Filter:** $Y_t = \int_{-\infty}^{\infty} w(t-s) X_s \, ds$, $t \in \mathcal{T}$. If $w(t) = 0$, for $t < 0$, then $t - s < 0$ implies that $w(t-s) = 0$. Thus $w(t-s) > 0$ only when $t - s > 0$, that is, when $t > s$. Then

$$Y_t = \int_{-\infty}^{\infty} w(t-s) X_s \, ds = \int_{-\infty}^{t} w(t-s) X_s \, ds.$$

Since we deal with values of the process in the past with respect to t, we say that the filter is *physically realizable.*

**General Physically Realizable Filter:** $Y_t = \int_{-\infty}^{t} w(t-s) X_s \, ds$, $t \in \mathcal{T}$.

The function, w(t), is called the *impulse response function*. We require $\int_{-\infty}^{\infty} |w(t)| \, dt < \infty$.

The domain, $\mathcal{T}$, of $X_t$ and $Y_t$, is, in general, a subset of $(-\infty, \infty)$. This subset may be $(-\infty, \infty)$, $(0, \infty)$, $\{0, \pm 1, \pm 2, \ldots\}$ for example. The backward shift, the difference and the moving average are examples of discrete-time filters. The exponential, differential, general linear filters and the general physically realizable filters are examples of continuous-time filters.

We may define physically realizable filters in the discrete-time case as well. For the filter, $Y_t = \sum_{j=-\infty}^{\infty} c_j X_{t-j}$, if $c_j = 0$ for $j < 0$, then $Y_t = \sum_{j=0}^{\infty} c_j X_{t-j}$ and $Y_t$ depends only on past values of $X_t$. This filter is physically realizable also.

In the discrete case we consider $g(\lambda) = \sum_{j=-\infty}^{\infty} c_j e^{-i\lambda j}$, for $|\lambda| \leq \pi$. In the continuous case, we let $g(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda t} w(t)\, dt$ $\lambda \in (-\infty, \infty)$. The function, $g(\lambda)$, is called the *transfer function* of the filter, while the function, $|g(\lambda)|^2$, is often called the
*frequency response function* of the filter. Notice that the transfer function is the Fourier transform of the impulse response function.

**8.2 Transfer Functions and Spectral Densities.** Now consider a stationary process $X_t$ with $E[X_t] = 0$ and spectral density $f_x(\lambda)$. The process $X_t$ is possibly complex valued. Let $Y_t = \sum_{j=-\infty}^{\infty} c_j X_{t-j}$ with $\sum_{j=-\infty}^{\infty} |c_j|^2 < \infty$
Then we have

$$E[Y_t] = E\left\{ \sum_{j=-\infty}^{\infty} c_j X_{t-j} \right\} = \sum_{j=-\infty}^{\infty} c_j E[X_{t-j}] = 0. \tag{8.2.1}$$

Thus the mean of a filtered process is 0 if the original process has mean 0. Consider the the expression for the covariance,

$$E[Y_t Y^*_{t+s}] = E\left\{ \sum_{j=-\infty}^{\infty} c_j X_{t-j} \sum_{k=-\infty}^{\infty} c^*_k X^*_{t+s-k} \right\}. \tag{8.2.2}$$

We may move the expectation operator through the summation signs to obtain

$$E[Y_t Y^*_{t+s}] = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_j c^*_k E[X_{t-j} X^*_{t+s-k}].$$

The expectation, however, is just the covariance of the $X_t$ process so that we have

$$E[Y_t Y^*_{t+s}] = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_j c^*_k \gamma_{s-k+j}, \tag{8.2.3}$$

which doesn't depend on t. Hence $Y_t$ is stationary stochastic process. This is true in general. A time-invariant linear filtered version of a stationary process is still stationary. We may continue the development of equation (8.2.3) by replacing the covariance function by the Fourier transform of the spectral density to obtain

$$E[Y_t Y^*_{t+s}] = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_j c^*_k \int_{-\pi}^{\pi} e^{-i\lambda(s-k+j)} f_x(\lambda)\, d\lambda. \tag{8.2.4}$$

We may reorder the integration and summations

$$E[Y_t Y^*_{t+s}] = \int_{-\pi}^{\pi} e^{-i\lambda s} \sum_{j=-\infty}^{\infty} c_j e^{-i\lambda j} \sum_{k=-\infty}^{\infty} c^*_k e^{+i\lambda k} f_x(\lambda)\, d\lambda$$

so that

$$E[Y_t Y^*_{t+s}] = \int_{-\pi}^{\pi} e^{-i\lambda s} \left| \sum_{j=-\infty}^{\infty} c_j e^{-i\lambda j} \right|^2 f_x(\lambda)\, d\lambda = \gamma^y_s. \tag{8.2.5}$$

But by (6.2.5), we know that $\gamma_s^y = \int_{-\pi}^{\pi} e^{-i\lambda s} f_y(\lambda)\, d\lambda$. Thus

$$\int_{-\pi}^{\pi} e^{-i\lambda s} \mid \sum_{j=-\infty}^{\infty} c_j\, e^{-i\lambda j} \mid^2 f_x(\lambda)\, d\lambda = \int_{-\pi}^{\pi} e^{-i\lambda s} f_y(\lambda)\, d\lambda. \tag{8.2.6}$$

By equating integrands in (8.2.6) we have

$$f_y(\lambda) = f_x(\lambda) \mid \sum_{j=-\infty}^{\infty} c_j\, e^{-i\lambda j} \mid^2 \tag{8.2.7}$$

or equivalently

$$f_y(\lambda) = f_x(\lambda) \mid g(\lambda) \mid^2. \tag{8.2.8}$$

Thus the spectral density of the $Y_t$ process may be computed in terms of the spectral density of the $X_t$ process that the transfer function of the filter. Frequently we may measure both the $X_t$ process and the $Y_t$ process. Thus by computing the ratio of the spectral densities, we may estimate the frequency response function of the filter.

The continuous time computation is similar. If $Y_t = \int_{-\infty}^{\infty} w(t-s) X_s ds$ with $\int_{-\infty}^{\infty} \mid w(t)\mid^2 dt < \infty$, we have

$$f_y(\lambda) = f_x(\lambda) \mid \int_{-\infty}^{\infty} w(x)\, e^{-i\lambda x}\, dx \mid^2 \tag{8.2.9}$$

so that

$$f_y(\lambda) = f_x(\lambda) \mid g(\lambda) \mid^2. \tag{8.2.10}$$

Thus we have shown the following theorem.

**Transfer Function Theorem.** If a zero mean stationary process $X_t$ with spectral density, $f_x(\lambda)$, is passed through a filter with transfer function, $g(\lambda)$, then

$$f_y(\lambda) = f_x(\lambda) \mid g(\lambda) \mid^2.$$

The nature of the frequency response function largely determines the properties of the filter. A filter such that

$$\mid g(\lambda) \mid^2 = \begin{cases} 1, & \lambda_1 \le |\lambda| \le \lambda_2 \\ 0, & \text{elsewhere} \end{cases} \tag{8.2.11}$$

is called a *band-pass filter.* A filter such that

$$\mid g(\lambda) \mid^2 = \begin{cases} 1, & |\lambda| \le \lambda_0 \\ 0, & \text{elsewhere} \end{cases} \tag{8.2.12}$$

is called a *low-pass filter.* Finally a filter such that

(8.2.13) $$| g(\lambda) |^2 = \begin{cases} 1, & |\lambda| \geq \lambda_0 \\ 0, & \text{elsewhere} \end{cases}$$

is called a *high-pass filter.*

Recall now that a white noise process, (that is a process which is a sequence of independent and identically distributed random variables), say $\epsilon_t$, has covariance $\gamma_s^\epsilon = 1$, for s $= 0$ and $\gamma_s^\epsilon = 0$, for s $\neq 0$. Thus the spectral density of a white noise process is

(8.2.14) $$f_\epsilon(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j^\epsilon \, e^{i\lambda j} = \frac{\gamma_0^\epsilon}{2\pi} = \frac{\sigma^2}{2\pi}$$

where $\sigma^2$ is the common variance of $\epsilon_t$. It is not too hard to think filtering the white noise with a time invariant linear filter to obtain

(8.2.15) $$X_t = \sum_{j=-\infty}^{\infty} c_j \, \epsilon_t.$$

By the transfer function theorem, $X_t$ has the spectral density function

(8.2.16) $$f_y(\lambda) = \frac{\sigma^2}{2\pi} \; | \sum_{j=-\infty}^{\infty} c_j \, e^{-i\lambda j} |^2.$$

This is a general linear model for a stochastic process known as an infinite moving average. If

(8.2.17) $$X_t = \sum_{j=0}^{q} c_j \, \epsilon_{t-j},$$

then

(8.2.18) $$f_y(\lambda) = \frac{\sigma^2}{2\pi} \; \left| \sum_{j=0}^{q} c_j \, e^{-i\lambda j} \right|^2.$$

The finite moving average representation of a stationary stochastic process, $X_t$, yields a powerful parametric representation in the time domain for a time series. Moreover we can see that the spectral density given in (8.2.18) has a parametric representation with $q + 1$ parameters. This motivates a general discussion of linear, time-domain models of time series with a finite number of parameters. These are known in the statistical literature as either Box-Jenkins models or autoregressive-moving average (ARMA) models and in the electrical engineering as linear recursive models. With this motivation, we shall begin our discussion of time domain modeling.

# Chapter 9
# Linear Processes

**9.1 Structural Models.** A structural model describes the relationship among two are more random variables in contrast to a statistical model which describes the probability structure of one or more random variables. To be sure there are very intimate connections which as we indicated earlier are captured in the time series case by the covariance structure of a stationary time series. Nonetheless, in the next chapters we are interested in describing structural models appropriate to time series. These time domain models are, in essence, analogs to linear regression models and, in fact, share so rather extensive theoretical connections. They do, however, also have features which are peculiar to time series analysis.

Let us first consider a discrete time process of the form

$$X_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \cdots$$

or in slightly more convenient form

$$X_t = \epsilon_t + \sum_{j=1}^{\infty} \psi_j \epsilon_{t-j}. \tag{9.1.1}$$

The $\epsilon_t$'s are uncorrelated random variables with zero means and constant variance, $\sigma_\epsilon^2$ . As defined earlier, such a series $\{\epsilon_j\}$ is called *white noise* and the process $\{X_t\}$ is called a *general linear process.* Notice that the covariance function for $\epsilon_t$ is given by

$$\gamma_k = \begin{cases} \sigma_\epsilon^2, & k=0 \\ 0, & k \neq 0 \end{cases} \tag{9.1.2}$$

and autocorrelation by

$$\rho_k = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0. \end{cases} \tag{9.1.3}$$

Equation (9.1.1) may be re-written under certain invertibility conditions which we describe later in the alternate form

$$X_t = \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots + \epsilon_t$$

or again in a slightly more compact form

$$X_t = \sum_{j=1}^{\infty} \pi_j X_{t-j} + \epsilon_t. \tag{9.1.4}$$

These infinite summation forms, (9.1.1) and (9.1.4), while elegant do not yield a convenient calculus for manipulating time series. It will be convenient to introduce difference equation notation for this will give us the tools for manipulating the time domain models more easily. Let *B* denote the backward shift operator so that

$BX_t = X_{t-1}$ and $B^jX_t = X_{t-j}$. Similarly $F$ is the forward shift operator so that $FX_t = X_{t+1}$ and, of course, $F^jX_t = X_{t+j}$. Thus we may write $X_t = (1 + \sum_{j=1}^{\infty} \psi_j B^j)\epsilon_t$ as

$$X_t = \psi(B)\epsilon_t \tag{9.1.5}$$

where, of course, the operator $\psi(B)$ is defined by $(1 + \sum_{j=1}^{\infty} \psi_j B^j)$. In a similar way, we may write $(1 - \sum_{j=1}^{\infty} \pi_j B^j)$ $X_t = \epsilon_t$ as

$$\pi(B)\, X_t = \epsilon_t. \tag{9.1.6}$$

Here the operator $\pi(B) = (1 - \sum_{j=1}^{\infty} \pi_j B^j)$. These operator forms of the infinite series representations of the time series allow us to develop a calculus. Consider applying $\psi(B)$ on both sides of (9.1.6) to obtain

$$\psi(B)\pi(B)\, X_t = \psi(B)\epsilon_t = X_t$$

from which we have the identity

$$\psi(B)\pi(B) = I. \tag{9.1.7}$$

Here $I$ is the identity operator so that $I\, X_t = X_t$. Notice also that $FB\, X_t = BF\, X_t = X_t$, so that $FB = BF = I$. It follows from this that $\psi(F)\psi(B) = I$. Combining this result with (9.1.7) we may conclude that $\psi(F) = \pi(B)$ or equivalently that $\psi^{-1}(B) = \pi(B)$ where inverse means the inverse operator.

**9.2 The Autocovariance Generating Function and the Spectral Density Function.** With the introduction of the notion of back shift and forward shift operators, we are now in a position to develop some simple difference equation theory. Let's first use (9.1.1) to notice that

$$E\,[X_tX_{t+\tau}] = E\left\{ \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi_j\psi_i\epsilon_{t-j}\epsilon_{t+\tau-i} \right\}.$$

We may interchange the expectation operator with the summation operators to obtain

$$E\,[X_tX_{t+\tau}] = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi_j\psi_i\, E[\epsilon_{t-j}\epsilon_{t+\tau-i}].$$

Because the $\epsilon_t$ process is white noise, only products with identical subscripts will contribute to the summation. Thus

$$E\,[X_tX_{t+\tau}] = \sum_{j=0}^{\infty} \psi_j\psi_{j+\tau}\, \sigma_\epsilon^2. \tag{9.2.1}$$

We first observe that the right hand side of (9.2.1) does not depend on t. Because $\epsilon_t$ is a 0 mean process, $X_t$ is also a 0 mean process. Thus (9.2.1) describes the covariance function $\gamma_t$ of a stationary process. We may thus write

$$E\,[X_tX_{t+\tau}] = \gamma_\tau = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j\psi_{j+\tau} \tag{9.2.2}$$

Our only concern is with the convergence of the infinite series represented in (9.2.2). We assume that $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Using Schwartz's inequality, we may conclude that $|\sum_{j=0}^{\infty} \psi_j\psi_{j+\tau}| < \infty$.

We may define $\gamma(s) = \sum_{\tau=-\infty}^{\infty} \gamma_\tau s^\tau$ to be the *autocovariance generating function.* We may substitute the expression (9.2.2) for $\gamma_\tau$ in the definition of the autocovariance generating function to obtain

$$\gamma(s) = \sigma_\epsilon^2 \sum_{\tau=-\infty}^{\infty} \sum_{j=0}^{\infty} \psi_j \psi_{j+\tau} s^\tau. \tag{9.2.3}$$

As is the usual routine, we interchange the order of summation

$$\gamma(s) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \sum_{\tau=-j}^{\infty} \psi_j \psi_{j+\tau} s^\tau$$

where it is understood that $\psi_j = 0$, $j < 0$. We may now re-index by letting $j + \tau = h$ so that $\tau = h - j$. With this we have

$$\gamma(s) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \sum_{h=0}^{\infty} \psi_j \psi_h s^{h-j}.$$

We may now separate the sums to obtain

$$\gamma(s) = \sigma_\epsilon^2 \left\{ \sum_{h=0}^{\infty} \psi_h s^h \right\} \left\{ \sum_{j=0}^{\infty} \psi_j s^{-j} \right\}.$$

Finally, re-writing in terms of $\psi(\cdot)$, we have

$$\gamma(s) = \sigma_\epsilon^2 \psi(s) \psi(s^{-1}). \tag{9.2.4}$$

As before, $B^{-1} B = I$ and also $FB = I$ so that $F = B^{-1}$. Thus substituting $B$ for s

$$\gamma(B) = \sigma_\epsilon^2 \psi(B) \psi(F) \tag{9.2.5}$$

As an example, consider a time series process whose structural equation is $X_t = \epsilon_t + \epsilon_{t-1}$. We may re-write this as $X_t = (1+B)\epsilon_t$. Thus $\psi(B) = I - B$ and $\gamma(B) = \sigma_\epsilon^2 (I+B)(I+B^{-1})$. Notice here we adopt the convention that $s^0 = 1$ is replaced with $B^0 = I$. We may expand this polynomial in $B$ to obtain $\gamma(B) = \sigma_\epsilon^2 (I+B+B^{-1}+BB^{-1})$ or $\gamma(B) = \sigma_\epsilon^2 (2\ I+B+B^{-1})$. We may use the autocovariance generating function to reconstruct the covariance function. Reading off the coefficients, we may deduce that $\gamma_0 = 2\sigma_\epsilon^2$, $\gamma_1 = \gamma_{-1} = \sigma_\epsilon^2$, and finally $\gamma_\tau = 0$, $|\tau| \geq 2$.

The autocovariance generating function may be used in another clever way. To see this application, let $s = e^{-i\lambda}$. Recall by definition, $\gamma(s) = \sum_{\tau=-\infty}^{\infty} \gamma_\tau s^\tau$. Thus when we substitute $e^{-i\lambda}$ for s, we obtain the spectral density function given below as

$$\gamma\left(e^{-i\lambda}\right) = \sum_{\tau=-\infty}^{\infty} \gamma_\tau e^{-i\lambda\tau} = 2\pi f(\lambda), \ -\pi \leq \lambda \leq \pi.$$

he covariance generating function is related to the spectral density function by the equation

$$f(\lambda) = \tfrac{1}{2\pi} \gamma(e^{-i\lambda}), \ -\pi \leq \lambda \leq \pi. \tag{9.2.6}$$

In our previous example, $X_t = \epsilon_t + \epsilon_{t-1}$, $\gamma(s) = \sigma_\epsilon^2 (2 + s + s^{-1})$. Thus

$$f(\lambda) = \tfrac{\gamma(e^{-i\lambda})}{2\pi} = \tfrac{\sigma_\epsilon^2}{2\pi} \left\{ 2+2\ \tfrac{(e^{-i\lambda}+e^{+i\lambda})}{2} \right\}, \ -\pi \leq \lambda \leq \pi.$$

We may simplify this to

$$f(\lambda) = \tfrac{\sigma_\epsilon^2}{2\pi}\ 2\ (1+\cos(\lambda)) = \tfrac{\sigma_\epsilon^2 (1+\cos(\lambda))}{\pi}, \ -\pi \leq \lambda \leq \pi$$

which is a particularly simple form of spectral density. Another informative example is the case for white noise, $X_t = \epsilon_t$. In this case $\gamma(s) = \sigma_\epsilon^2$ so that $f(\lambda) = \tfrac{\sigma_\epsilon^2}{2\pi}$, a constant for $-\pi \leq \lambda \leq \pi$. This gives insight to the use of the phrase, white noise. As with white light, white noise contains all frequencies with equal amplitude. The fact that the spectral density function is constant verifies that all frequencies are of equal amplitude.

The general spectral density is computed by first observing that the autocovariance generating function is given by

$$\gamma(s) = \sigma_\epsilon^2 \psi(s) \psi(s^{-1}).$$

Substituting $e^{-i\lambda}$ for s, we have

$$f(\lambda) = \tfrac{\sigma_\epsilon^2}{2\pi} \psi(e^{-i\lambda}) \psi(e^{i\lambda}) = \tfrac{\sigma_\epsilon^2}{2\pi} |\psi(e^{-i\lambda})|^2. \tag{9.2.7}$$

Thus if we know the infinite moving average form of the general linear model, then we may calculate the spectral density function directly by (9.2.7). This is a very powerful equation as it allows us to move from a time domain model to a frequency domain model easily. It also has the enormous advantage that if we have a

finite parameter time domain model, we may easily calculate the spectral density parametrically in stead of having to use the nonparametric smoothing windows techniques we studied in Chapter 7.

**9.3 Stationarity and Invertibility.** We have already determined that for the infinite moving average representation, $X_t$ is always stationary provided that $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ Suppose we assume that $\sum_{j=0}^{\infty} \psi_j < \infty$. For this series to be convergent, $\psi_j \rightarrow 0$ so that $|\psi_j| \rightarrow 0$ as well. Thus eventually $|\psi_j| < 1$, so that $|\psi_j|^2 \leq |\psi_j|$. This implies that $\sum_{j=0}^{\infty} |\psi_j|^2 < \infty$ so that

$$\psi(s) = \sum_{j=0}^{\infty} \psi_j s^j$$

exists on and within the unit circle, $|s| \leq 1$.

If a time series process with $\psi$-weights can be changed to a process with converging $\pi$-weights, it is said to be *invertible*. A series may not be invertible because the $\pi$ weights increase instead of decrease. Consider for example $X_t = (I - \psi B)\epsilon_t$. The inverse operator is $(I - \psi B)^{-1} = (I + \psi B + \psi^2 B^2 + \cdots)$ so that $X_t = \epsilon_t - \psi X_{t-1} - \psi^2 X_{t-2} - \cdots$.
Thus $\pi_j = -\psi^j$ so if $|\psi|>1$, the $\pi$ weights increase. We say $X_t$ is *invertible* if the generating function of the $\pi$ weights,

$$\pi(s) = \sum_{j=0}^{\infty} \pi_j s^j$$

exists on and within the unit circle, $|s| \leq 1$.

In the special case of the $\pi$-weight representation when we have only a finite number of $\pi$ weights, we call the process *autoregressive* and write

(9.3.1) $$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t.$$

We denote such a process by AR(p) and p is called the *order* of the process. We can write $(I - \phi_1 B - \cdots - \phi_p B^p) X_t = \epsilon_t$ or in operator form

(9.3.2) $$\phi(B) X_t = \epsilon_t$$

where, of course, $\phi(B) = (I - \phi_1 B - \cdots - \phi_p B^p)$.

If only a finite number of the $\psi$ weights are non-zero, we call the process a *moving average process* and write

(9.3.3) $$X_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q}.$$

We denote such a process by MA(q) where now q is called the *order* of the process. We can write $X_t = (I - \theta_1 B - \cdots - \theta_q B^q)\epsilon_t$ or in operator form

(9.3.4) $$X_t = \theta(B)\epsilon_t.$$

If we have more complicated data, it is sometimes necessary to use a mixed model which is given in operator form by

(9.3.5) $$\phi(B)X_t = \theta(B)\epsilon_t.$$

This is called a *mixed autoregressive moving average* process of order (p, q) and abbreviated ARMA (p, q). These three classes of models represent the most general form of finite parameter linear models for the structure of a time series process and are commonly know as *Box-Jenkins models*. We will study these time domain models in more detail. We note in closing this section that autoregressive processes are always invertible by definition and that moving average processes are always stationary. Autoregressive and ARMA processes may be stationary and moving average and ARMA processes may be invertible. Conditions under which they are are the subject of our continued study.

# Chapter 10
# Autogregressive Models

**10.1 Autoregressive Models.** Let us now consider a general $p^{th}$ order autoregressive model, $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t$ or $(I - \phi_1 B - \cdots - \phi_p B^p)\ X_t = \phi(B)\ X_t = \epsilon_t$. Clearly this is an invertible process, so the question in this case is whether it is stationarity. For stationarity $\psi(s)$ must converge for $|s| \leq 1$. We know by definition that $\phi(s)$ is a $p^{th}$ order polynomial. Suppose we write $\phi$ in terms of its factors

$$\phi(s) = (I - u_1 s)(I - u_2 s) \cdots (I - u_p s). \tag{10.1.1}$$

We may expand (10.1.1) in terms of partial fractions as follows

$$\frac{1}{\phi(s)} = \frac{1}{(1-u_1 s)\cdots(1-u_p s)} = \frac{k_1}{(1-u_1 s)} + \cdots + \frac{k^p}{(1-u_p s)} \tag{10.1.2}$$

so that

$$X_t = \phi^{-1}(B)\ \epsilon_t = \sum_{i=1}^{p} \frac{k_i}{1-u_i B}\ \epsilon_t.$$

Now expand $\frac{1}{1-u_i B}\ \epsilon_t$ as $(1 + u_i B + u_i^2 B^2 + \cdots)\ \epsilon_t$ which is stationary if $|u_i| < 1$. Hence the $p^{th}$ order autoregressive process is stationary if $|u_i| < 1$ for i = 1, 2, ..., p. Setting $\phi(s) = 0$, we find $(1 - u_1 s) \cdots (1 - u_p s) = 0$. Hence the roots are the polynomial $\phi(s)$ are $s = \frac{1}{u_1}, \ldots, \frac{1}{u_p}$. Thus $|\frac{1}{u_i}|>1$ if and only if $|u_i| < 1$, so that an autoregressive process is stationary if and only if the roots of the equation, $\phi(s) = 0$, lie outside the unit circle. The equation, $\phi(s) = 0$, is called the *auxiliary equation* and has much practical importance in time series analysis.

Let us again consider the autogressive model, $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t$. We wish to develop an expression for the autocovariance function. Multiplying through this expression by $X_{t-\tau}$, we obtain

$$X_t X_{t-\tau} = \phi_1 X_{t-1} X_{t-\tau} + \cdots + \phi_p X_{t-p} X_{t-\tau} + \epsilon_t X_{t-\tau}\ . \tag{10.1.3}$$

Taking expectations, we see that

$$\gamma_\tau = \phi_1 \gamma_{\tau-1} + \cdots + \phi_p \gamma_{\tau-p},\ \tau > 0. \tag{10.1.4}$$

Note that $X_{t-\tau}$ depends only on $\epsilon_{t-\tau}, \epsilon_{t-\tau-1}, \ldots$ and so is uncorrelated with $\epsilon_t$. Thus $E[\epsilon_t X_{t-\tau}] = 0$. Dividing (10.1.4) by $\gamma_0$, we obtain an expression for the autocorrelation function

$$\rho_\tau = \phi_1 \rho_{\tau-1} + \cdots + \phi_p \rho_{\tau-p,\ \tau>0,}$$

or equivalently

$$\phi(B)\rho_\tau = 0. \tag{10.1.5}$$

This latter expression is a *homogeneous linear difference equation* whose solution is given as a linear combination of a function of the roots of the auxiliary equation.

In particular if $\phi(s) = (1 - u_1 s)\cdots(1 - u_p s)$ and all the $u_i$ are distinct, then

$$\rho_\tau = A_1 u_1^\tau + A_2 u_2^\tau + \cdots + A_p u_p^\tau \tag{10.1.6}$$

where $\tau$ is an integer. Consider, for example, the term $A_1 u_1^\tau$. Operate on it with the operator, $(I - u_1 B)$. Then $(I - u_1 B)\ A_1\ u_1^\tau = A_1 u_1^\tau - u_1 A_1 u_1^{\tau-1} = 0$. Similarly the term $A_j u_j^\tau$ is eliminated by $(I - u_j B)$. If $u_j^{-1}$ is a double root, i.e. if $\phi(s)$ contains $(1 - u_j s)^2$, then the solution becomes $A_j u_j^\tau + A_{j+1} \tau u_j^\tau$. Note that $(I - u_j B)^2\ A_{j+1} \tau u_j^\tau = (I - u_j B)[A_{j+1} \tau u_j^\tau - u_j A_{j+1}(\tau - 1) u_j^{\tau-1}] = (I - u_j B)[A_{j+1} u_j^\tau] = A_{j+1} u_j^\tau - u_j A_{j+1} u_j^{\tau-1} = 0$. Similarly for higher order multiple roots.

Now returning to $\rho_\tau = \phi_1 \rho_{\tau-1} + \cdots + \phi_p\ \rho_{\tau-p}$, noting the symmetry, $\rho_{-\tau} = \rho_\tau$, and letting k = 1, 2, ..., p, we obtain a series of equations

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \cdots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \cdots + \phi_p \rho_{p-2} \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \cdots + \phi_p \end{aligned} \tag{10.1.7}$$

This set of equations is known as the *Yule-Walker* equations and they relate the autoregressive coefficients to the autocorrelations (or equivalently the autocovariances). In matrix form write

$$\boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_p \end{bmatrix}, \boldsymbol{\rho} = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_p \end{bmatrix}, \mathbf{P_p} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & & & & \vdots \\ \rho_{p-1} & \rho_{p-2} & & \cdots & 1 \end{bmatrix}$$

So that the Yule-Walker equations become

(10.1.8) $$\boldsymbol{\phi} = \mathbf{P}_{\boldsymbol{p}}^{-1}\boldsymbol{\rho}.$$

We have used the basic autoregressive difference equation to derive the $\phi_j$ coefficients in terms of the autocorrelations $\rho_\tau$. Next we wish to derive the variance, $\sigma_\epsilon^2$, in terms of the autocorrelations and the autoregressive coefficients. Let's consider again $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t$. Multiplying by $X_t$, we obtain

$$X_t^2 = \phi_1 X_{t-1} X_t + \cdots + \phi_p X_{t-p} X_t + \epsilon_t X_t.$$

Noting that $E[\epsilon_t X_t] = E[\epsilon_t\ \psi(B)\epsilon_t] = E[\epsilon_t^2] = \sigma_\epsilon^2$, we obtain

(10.1.9) $$\gamma_0 = \phi_1\gamma_1 + \cdots + \phi_p\gamma_p + \sigma_\epsilon^2.$$

Now dividing by $\gamma_0 = \sigma_z^2$,

$$1 = \phi_1\rho_1 + \cdots + \phi_p\rho_p + \frac{\sigma_\epsilon^2}{\sigma_z^2}$$

or

$$\sigma_z^2 = \frac{\sigma_\epsilon^2}{1 - \phi_1\rho_1 - \cdots - \phi_p\rho_p}.$$

Also noting $\psi(s) = \frac{1}{\phi(s)}$, the spectral density becomes

$$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi}|\psi(e^{i\lambda})|^2 = \frac{\sigma_\epsilon^2}{2\pi}\frac{1}{|1 - \phi_1 e^{-i\lambda} - \cdots - \phi_p e^{-ip\lambda}|^2}.$$

**10.2. First Order Autoregressive Process (Gauss-Markov Process).** The first order autoregressive process is $X_t = \phi_1 X_{t-1} + \epsilon_t = \epsilon_t + \phi_1\epsilon_{t-1} + \phi_1^2\epsilon_{t-2} + \cdots$ where $-1 < \phi_1 < 1$ for stationarity. The auxiliary equation is $\phi(s) = (1 - \phi_1 s) = 0$, so that the difference equation $\rho_\tau = \phi_1\rho_{\tau-1}$ has solution $\rho_\tau = \phi_1^\tau$, $k \geq 0$. In particular $\rho = \rho_1 = \phi_1$ so that often the process is written $X_t = \rho X_{t-1} + \epsilon_t$. The autocorrelation function either decays exponentially $\phi_1 \geq 0$ or oscillates, but the oscillations decay exponentially. The variance $\sigma_z^2$ is given by

(10.2.1) $$\sigma_z^2 = \frac{\sigma_\epsilon^2}{1 - \rho_1\phi_1} = \frac{\sigma_\epsilon^2}{1 - \rho_1^2}.$$

The spectral density $f(\lambda)$ is given by

(10.2.2) $$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi}\frac{1}{|1 - \phi_1 e^{-i\lambda}|^2}, \quad -\pi \leq \lambda \leq \pi.$$

This may be expanded as

$$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi}\frac{1}{1 + \phi_1^2 - \phi_1(e^{-i\lambda} + e^{+i\lambda})}, \quad -\pi \leq \lambda < \pi$$

or

$$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \frac{1}{1+\phi_1^2-2\phi_1\cos(\lambda)}.$$

**10.3 Second Order Autoregressive Process.** The second order autoregressive process is

(10.3.1) $$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t.$$

For stationarity, the roots of the auxiliary equation, $\phi(s) = 1 - \phi_1 s - \phi_2 s^2 = 0$, must be outside the unit circle. A necessary and sufficient condition for this is that *i.* $\phi_1+\phi_2 < 1$, *ii.* $\phi_2 - \phi_1 < 1$, and *iii.* $-1 < \phi_2 < 1$. This is a triangular region of admissibility for the coefficients $\phi_1$ and $\phi_2$. **** Assuming distinct roots, the solution to the difference equation, $\rho_\tau = \phi_1\rho_{\tau-1} + \phi_2\rho_{\tau-2}$, is

(10.3.2) $$\rho_\tau = A_1 u_1^\tau + A_2 u_2^\tau.$$

We need some initial conditions to solve for the unknown coefficients, $A_1$ and $A_2$. Recall that $\rho_0 = 1$ and, from the Yule-Walker equations, we have $\rho_1 = \phi_1+\phi_2\rho_1$. Hence $\rho_1 = \frac{\phi_1}{1-\phi_2}$. Now since $1 - \phi_1 s - \phi_2 s^2 = (1 - u_1 s)(1 - u_2 s) = 1 - (u_1+u_2)s + u_1u_2s^2$, we can write $\phi_1 = u_1+u_2$ and $\phi_2 = -u_1u_2$. Thus

$$\rho_1 = \frac{u_1+u_2}{1+u_1u_2}.$$

Using this together with the condition that $\rho_0 = 1$, we can write

(10.3.3) $$\rho_0 = A_1+A_2 = 1 \text{ and } \rho_1 = A_1u_1+A_2u_2 = \frac{u_1+u_2}{1+u_1u_2}.$$

Note the symmetry in $u_1$ and $u_2$, $A_1$ and $A_2$. Now

$$A_1u_1 + (1 - A_1)u_2 = \frac{u_1+u_2}{1+u_1u_2}$$

which implies

$$A_1(u_1 - u_2) = \frac{u_1+u_2-u_2-u_1u_2^2}{1+u_1u_2}$$

so that

(10.3.4) $$A_1 = \frac{u_1(1-u_2^2)}{(u_1-u_2)(1+u_1u_2)}.$$

By symmetry,

(10.3.5) $$A_2 = \frac{u_2(1-u_1^2)}{(u_2-u_1)(1+u_1u_2),}$$

hence

(10.3.6) $$\rho_\tau = \frac{(1-u_2^2)u_1^{\tau+1} - (1-u_1^2)u_2^{\tau+1}}{(u_1-u_2)(1+u_1u_2)}.$$

In this case $\rho_\tau$ is a mixture of exponentially decaying functions possibly oscillating if either of $u_1$ or $u_2$ is negative.

If the roots are real and equal, say $u^{-1}$, then $\rho_\tau = A_1u^\tau+A_2\tau u^\tau$. Again using the initial conditions, $\rho_0 = A_1 = 1$ and $\rho_1 = A_1u+A_2u = \frac{2u}{1+u^2}$. Thus

$$A_2 + 1 = \frac{2}{1+u^2}$$

so that

(10.3.7) $$A_2 = \frac{1-u^2}{1+u^2}.$$

Hence

$$\rho_\tau = u^\tau+\frac{1-u^2}{1+u^2}\tau u^\tau = \frac{1+u^2+\tau-\tau u^2}{1+u^2}\, u^\tau.$$

Thus

(10.3.8) $$\rho_\tau = \frac{1+\tau+(1-\tau)u^2}{1+u^2}\, u^\tau.$$

Thus $\rho_\tau$ either decays exponentially or is dominated by a decaying exponential.

Finally, if the roots are complex conjugate pairs, we can write $u_1 = ue^{i\lambda_0}$ and $u_2 = ue^{-i\lambda_0}$. Since $|u_1| < 1$, we also have $|u| < 1$. Now

$$\rho_\tau = A_1u^\tau e^{i\lambda_0\tau} + A_2u^\tau e^{-i\lambda_0\tau}.$$

Taking complex conjugates,

$$\rho_\tau^* = \rho_\tau = A_1^* u^\tau e^{-i\lambda_0\tau} + A_2^*u^\tau e^{i\lambda_0\tau}.$$

Since $\rho_\tau$ is real, we can identify $A_1^* = A_2$ and $A_2^* = A_1$. Let $A_1 = \frac{1}{2}(a - bi)$ and $A_2 = \frac{1}{2}(a + bi)$. Then

$$\rho_\tau = \tfrac{1}{2}(a - bi)u^\tau e^{i\lambda_0\tau}+\tfrac{1}{2}(a+bi)u^\tau e^{-i\lambda_0\tau}.$$

Collecting coefficients, we have

(10.3.9) $$\rho_\tau = u^T \left\{ a\frac{(e^{i\lambda_0\tau}+e^{-i\lambda_0\tau})}{2} + b\frac{(e^{i\lambda_0\tau}-e^{-i\lambda_0\tau})}{2i} \right\}.$$

Using our earlier trigonometric identities,

(10.3.10) $$\rho_\tau = u^T \Big\{ a\cos(\lambda_0\tau) + b\sin(\lambda_0\tau) \Big\}$$

where a and b can be written in terms of $A_1$. Once again we find $\rho_\tau$ is dominated by a decaying exponential.

Using the Yule-Walker Equations, $\rho_1 = \phi_1 + \phi_2\rho_1$ and $\rho_2 = \phi_1\rho_1 + \phi_2$, we may obtain

(10.3.11) $$\rho_1^2 = \phi_1\rho_1 + \phi_2\rho_1^2 \text{ and } \rho_2 = \phi_1\rho_1 + \phi_2$$

may be combined to solve for the $\phi_2$

(10.3.12) $$\phi_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}.$$

From this $\phi_1 = (1-\phi_2)\rho_1 = \frac{1-\rho_1^2-(\rho_2-\rho_1^2)}{1-\rho_1^2}\rho_1$, so that

(10.3.13) $$\phi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}.$$

Solving in turn for $\rho_1$ and $\rho_2$,
(10.3.14) $\rho_1 = \phi_1 1_{1-\phi_2}$

and

(10.3.15) $$\rho_2 = \frac{\phi_1^2}{1-\phi_2} + \phi_2$$

The variance $\sigma_z^2$ is given by

$$\sigma_z^2 = \frac{\sigma_\epsilon^2}{1-\rho_1\phi_1-\rho_2\phi_2}.$$

Substituting (10.3.14) and (10.3.15) in this expression

$$\sigma_z^2 = \frac{\sigma_\epsilon^2}{1-\frac{\phi_1^2}{1-\phi_2}-\left\{\phi_2+\frac{\phi_1^2}{1-\phi_2}\right\}\phi_2}.$$

Simplifying by multiplying top and bottom by $(1-\phi_2)$,

$$\sigma_z^2 = \frac{(1-\phi_2)\sigma_\epsilon^2}{1-\phi_2-\phi_1^2-\phi_2^2(1-\phi_2)-\phi_1^2\phi_2}.$$

Now collecting coefficients

(10.3.16) $$\sigma_z^2 = \frac{1-\phi_2}{1+\phi_2}\,\frac{\sigma_\epsilon^2}{(1-\phi_2)^2-\phi_1^2}.$$

Finally the spectral density is given by

(10.3.17) $$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi}\,\frac{1}{|1-\phi_1 e^{-i\lambda}-\phi_2 e^{-2i\lambda}|^2}.$$

Expanding the expression in the denominator, we obtain

$$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \frac{1}{1+\phi_1^2+\phi_2^2+2\phi_1(\phi_2-1)\left\{\frac{e^{i\lambda}+e^{-i\lambda}}{2}\right\} - 2\phi_2\left\{\frac{e^{2i\lambda}+e^{-2i\lambda}}{2}\right\}}.$$

Using trigonometric identities once again, we obtain the closed form for the second order autoregressive spectral density

(10.3.18) $$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \frac{1}{1+\phi_1^2+\phi_2^2-2\phi_1(1-\phi_2)\cos(\lambda) - 2\phi_2\cos(2\lambda)}.$$

Thus the theory of autoregressive processes is quite complete with respect to expressing the autoregressive parameters $\phi_1,\ldots,\phi_p$, $\sigma_\epsilon^2$ and, by extension, f($\lambda$), in terms of the autocovariance, $\gamma_\tau$. Since we have the ability to estimate $\gamma_\tau$ directly from the data, we have a reasonably satisfactory approach to estimating the autoregressive structure in general. Two facts are evident in this discussion. First, we are estimating functions of the moments, that is, functions of the covariances. This is essentially applying the method of moments, which, although it has a long history, is certainly not guaranteed to give us any statistical optimality. It would be better to consider least squares or maximum likelihood techniques if available. We shall do this later. Also, left unsaid was how to determine the order, p, of the process. We address this issue by the use of partial autocorrelations.

**10.4 Partial Autocorrelations.** The autocorrelation, $\rho_\tau$, is a valuable diagnostic tool in fitting time series models. As we have just seen, $\rho_\tau$ is a function which either is a decreasing exponential function or at least is dominated by a decreasing exponential function whenever $X_t$ is an autoregressive process. A second function which will serve as a diagnostic aid is the *partial autocorrelation function of lag k*, $\phi_{\tau\tau}$. To define this let $\phi_{\tau j}$ be the $j^{th}$ coefficient of an autoregressive process of order $\tau$. Thus, our standard $p^{th}$ order autoregressive model becomes

$$\rho_j = \phi_{\tau 1}\rho_{j-1} + \phi_{\tau 2}\rho_{j-2} + \cdots + \phi_{\tau\tau}\rho_{j-\tau} \quad j = 1, 2, \ldots, \tau.$$

From the Yale-Walker equations, we have

(10.4.1) $$\begin{bmatrix} \phi_{\tau 1} \\ \vdots \\ \phi_{\tau\tau} \end{bmatrix} = \mathbf{P}_{\boldsymbol{\tau}}^{-1}\boldsymbol{\rho}_{\boldsymbol{\tau}}$$

so that $\phi_{\tau\tau}$ can be easily picked out. Note $\phi_{11} = \rho_1$. Also worth noting is that if $X_t$ is truly an autoregressive process of order p, the $\sigma_{\tau\tau}$ will be non-zero for $\tau \leq$ p.

The partial autocorrelations may be estimated by fitting autoregressive processes of orders 1, 2, 3, ... and picking out the estimates $\hat{\phi}_{11}$, $\hat{\phi}_{22}$, $\hat{\phi}_{33}$, .... If $X_t$ is an autoregressive process of order p, then for $\tau \geq$ p+1, $\hat{\phi}_{\tau\tau}$ should approximate 0 and the its variance given by

$$\text{var}\,[\hat{\phi}_{\tau\tau}] \simeq \frac{1}{n}$$

so that the standard error of $\hat{\phi}_{\tau\tau}$ is approximated by

(10.4.2) $$\text{SE}[\hat{\phi}_{\tau\tau}] \simeq \frac{1}{\sqrt{n}}, \tau \geq \text{p+1}.$$

As a diagnostic tool, we can thus plot the $\hat{\phi}_{\tau\tau}$ as a function of $\tau$. By drawing in the SE limits we can compare the values of $\hat{\phi}_{\tau\tau}$ to the standard error estimates. If $\hat{\phi}_{\tau\tau}$ lies well within the two standard errors limit, it will probably be estimating $\phi_{\tau\tau} = 0$ and hence the process is of order less than $\tau$.

# Chapter 11
# Moving Average Processes

**11.1 Moving Average Processes** We now turn our attention to a moving average process of order q given by the expression, $X_t = \epsilon_t - \theta_1\epsilon_{t-1} - \cdots - \theta_q\epsilon_{t-q}$. Expressing thus in terms of a backwards shift operator, we obtain

$$X_t = (I - \theta_1 B - \cdots - \theta_q B^q)\epsilon_t = \theta(B)\epsilon_t \tag{11.1.1}$$

Recall that a moving average process is always stationary, but may not be invertible. However, in its inverted form we would have $\epsilon_t = \theta^{-1}(B)\, X_t$. If we let

$$\theta(s) = \prod_{j=1}^{q}(1 - v_j s)$$

and expanding as we did in the autoregressive case using partial fractions, we obtain

$$\pi(s) = \theta^{-1}(s) = \sum_{j=1}^{q} \frac{m_j}{1 - v_j B}.$$

As with the stationary conditions for an autoregressive process,

$$\frac{1}{1 - v_j B} = 1 + v_j B + v_j^2 B^2 + \cdots$$

provide $|v_j| < 1$. Hence $X_t$ will be *invertible* if $|v_j| < 1$ for each $j = 1, 2, \ldots, q$. Or equivalently $X_t = \theta(B)\epsilon_t$ is invertible if and only if the roots of this auxiliary equation, $\theta(s) = 0$, lie outside the unit circle.

Just as the autoregressive process was always invertible because $|\pi(s)| < \infty$ for all s, so too is the moving average process always stationary because $|\psi(s)| = |\theta(s)| < \infty$ for all s. It is worthwhile to note this duality between moving average and autoregressive processes. It recurs frequently. To compute the autocorrelation function for a MA(q) process, consider

$$\gamma_\tau = E[X_t X_{t-\tau}] = E[(\epsilon_t - \theta_1\epsilon_{t-1} - \cdots - \theta_q\epsilon_{t-q})(\epsilon_{t-\tau} - \theta_1\epsilon_{t-\tau-1} - \cdots - \theta_q\epsilon_{t-\tau-q})]$$

Hence, since only terms with similar subscripts on the $\epsilon$ will contribute to the covariance, we have

$$\gamma_0 = (1 + \theta_1^2 + \cdots + \theta_q^2)\,\sigma_\epsilon^2 \tag{11.1.2}$$

and

$$\gamma_\tau = \begin{cases} (-\theta_\tau + \theta_1\theta_{\tau+1} + \cdots + \theta_{q-\tau}\theta_q)\sigma_\epsilon^2, & \tau = 1, 2, \ldots, q \\ 0, & \tau \geq q+1 \end{cases}. \tag{11.1.3}$$

Thus the autocorrelation is

$$\rho_\tau = \begin{cases} \dfrac{-\theta_\tau + \theta_1\theta_{\tau+1} + \cdots + \theta_{q-\tau}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2}, & \tau = 1, 2, \ldots, q \\ 0, & \tau \geq q+1. \end{cases} \tag{11.1.4}$$

Notice for the MA(q) process, the autocorrelations behave like the partial auto-correlations do for the AR(p) in the sense that they are non-zero up to some cut-off point q and the zero from then on.

Similarly for the moving average process the partial autocorrelations will be dominated by an exponential function. Again the duality between autoregressive and moving average shows up. Comparison of these two functions, the autocorrelation and the partial autocorrelation, frequently allows us to distinguish between data which follows autoregressive and moving average processes. Since $\psi(s) = \theta(s) = 1 - \theta_1 s - \cdots - \theta_q s^q$, the spectral density function is

$$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} |\psi(e^{-i\lambda})|^2 = \frac{\sigma_\epsilon^2}{2\pi} |1 - \theta_1 e^{-i\lambda} - \cdots - \theta_q e^{-qi\lambda}|^2. \tag{11.1.5}$$

**11.2 The First Order Moving Average Process** It is instructive to consider several simple cases of the moving average process. We begin with the first order moving average, $X_t = \epsilon_t - \theta_1 \epsilon_{t-1}$. In the operator form we have,

$$X_t = (I - \theta_1 B)\epsilon_t = \theta(B)\epsilon_t. \tag{11.2.1}$$

For invertibility, $-1 < \theta_1 < 1$ since $\frac{1}{\theta_1}$ is the root of the auxiliary equation $\theta(s) = 0$. The variance is given by

$$\gamma_0 = (1+\theta_1^2)\sigma_\epsilon^2 \tag{11.2.2}$$

and the autocorrelation function

$$\rho_\tau = \begin{cases} \frac{-\theta_1}{1+\theta_1^2}, & \tau = 1 \\ \\ 0, & \tau \geq 2 \end{cases}. \tag{11.2.3}$$

Of course, as always $\rho_0 = 1$. We note from this equation, $\rho_1 = (-\theta_1)/(1+\theta_1^2)$, that $\theta_1$ and $\rho_1$ are related in a non-linear fashion. In fact if $\theta_1$ is a solution then so is $\theta_1^{-1}$ since

$$\rho_1 = \frac{-\frac{1}{\theta_1}}{1+\frac{1}{\theta_1^2}} = \frac{-\theta_1}{\theta_1^2+1} = \frac{-\theta_1}{1+\theta_1^2}.$$

The spectral density function for the first order moving average process is

$$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} |1 - \theta_1 e^{-i\lambda}|^2. \tag{11.2.4}$$

Expanding the square modulus

$$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \left\{ 1+\theta_1^2 - 2\theta_1 \left( \frac{e^{-i\lambda}+e^{i\lambda}}{2} \right) \right\}$$

which reduces to

$$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \left\{ 1 + \theta_1^2 - 2\theta_1 \cos(\lambda) \right\}. \tag{11.2.5}$$

We can calculate the partial autocorrelations from the Yule-Walker equations and our expressions for $\rho_\tau$. In particular

$$\mathbf{P}_\tau = \begin{bmatrix} 1 & \frac{-\theta_1}{1+\theta_1^2} & 0 & \dots & 0 & 0 \\ \frac{-\theta_1}{1+\theta_1^2} & 1 & \frac{-\theta_1}{1+\theta_1^2} & \cdots & 0 & 0 \\ \vdots & & \vdots & & & \vdots \\ 0 & 0 & 0 & \cdots & \frac{-\theta_1}{1+\theta_1^2} & 1 \end{bmatrix}$$

so that

$$\mathbf{P}_\tau \begin{bmatrix} \phi_{\tau 1} \\ \vdots \\ \phi_{\tau\tau} \end{bmatrix} = \begin{bmatrix} \frac{-\theta_1}{1+\theta_1^2} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Solving this for $\theta_{\tau\tau}$ yields,

(11.2.6) $$\hat{\phi}_{\tau\tau} = -\theta_1^\tau[1-\theta_1^2]/[1-\theta_1^{2(\tau+1)}]$$

which is dominated by an decreasing exponential. Notice that the partial autocorrelation for the moving average process has this damped exponential property that the autocorrelation of the autoregressive process has and simultaneously the autocorrelation of the moving average process has the cutoff property that the partial autocorrelation of the autoregressive process has.

**11.3 The Second Order Moving Average Process** The second order moving average process is $X_t = \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2}$. Of course it is automatically stationary. In analogy with the AR(2) process, it will be invertible if and only if the roots of the equation, $1 - \theta_1 s - \theta_2 s^2 = 0$, lie outside the unit circle. Equivalently, again in analogy to the stationary conditions for the AR(2) case, if and only if *i.* $\theta_2+\theta_1 < 1$, *ii.* $\theta_2 - \theta_1 < 1$ and *iii.* $-1 < \theta_2 < 1$. The variance of the second order moving average process is

(11.3.1) $$\gamma_0 = \sigma_\epsilon^2(1+\theta_1^2+\theta_2^2)$$

and the autocorrelation

$$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$$

(11.3.2) $$\rho_2 \;=\; \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$$

and

$$\rho_\tau = \; 0, \tau \;\geq\; 3.$$

The spectral density

(11.3.3) $$\mathrm{f}(\lambda) \;=\; \frac{\sigma_\epsilon^2}{2\pi} \,|1 - \theta_1 \mathrm{e}^{-i\lambda} - \theta_2 \mathrm{e}^{-2i\lambda} \,|^{\,2}$$

which may be written in real terms as

(11.3.4) $$\mathrm{f}(\lambda) \;=\; \frac{\sigma_\epsilon^2}{2\pi} \left\{1+\theta_1^2+\theta_2^2 - 2\,\theta_1(1-\theta_2)\cos(\lambda) - 2\theta_2\cos(2\lambda)\right\}, \; -\pi < \lambda < \pi.$$

Again it is clear that the basic theory for moving average time series is in hand. The major difficulty is that, in general, the relationship between the moving average coefficients, $\theta_j$, and the covariance function, $\gamma_\tau$, is nonlinear whereas the relationship between the autoregressive coefficients, $\phi_j$, and the covariance function, $\gamma_\tau$, is given in a linear fashion by the Yule-Walker equations. For this reason as well as their analogy to differential equations modeling physical systems, the autoregressive models seem to be somewhat preferred. Nonetheless, the autoregressive models may not be a very economical representation. The most general finite parameter representation of a linear process is given by the mixed autoregressive-moving average models which we study in the next section.

**11.4 Mixed Processes-Autoregressive-Moving Average Processes** In order to achieve a representation of a time series data set with a finite number of parameters, we can consider a mixed model

(11.4.1) $$\mathrm{X}_t \;=\; \phi_1 \mathrm{X}_{t-1} + \cdots + \phi_p \mathrm{X}_{t-p} + \epsilon_t \,-\, \theta_1 \epsilon_{t-1} \,-\, \cdots \,-\, \theta_q \epsilon_{t-q}$$

or rewriting in operator form

(11.4.2) $$\phi(B)\mathrm{X}_t \;=\; \theta(B)\epsilon_t.$$

We refer to this as an ARMA (p, q) process. Noting that

$$\mathrm{X}_t \;=\; \phi^{-1}(B)\theta(B)\epsilon_t$$

we infer that the process will be stationary if $\phi^{-1}(B)$ exists, that is, if the roots of the auxiliary equation, $\phi(\mathrm{s}) = 0$, lie outside the unit circle. By the same token,

$$\theta^{-1}(B)\phi(B)\;\mathrm{X}_t \;=\; \epsilon_t$$

so that the process is invertible if $\theta^{-1}(B)$ exists, that is, if the roots of the auxiliary equation, $\theta(\mathrm{s}) = 0$, lie outside the unit circle.

Next, we multiply (11.4.1) by $\mathrm{X}_{t-\tau}$ to obtain

(11.4.3) $$\begin{aligned}\mathrm{X}_t\mathrm{X}_{t-\tau} \;=\; & \phi_1 \mathrm{X}_{t-1}\mathrm{X}_{t-\tau} + \cdots + \phi_p \mathrm{X}_{t-p}\mathrm{X}_{t-\tau} \\ & + \epsilon_t \mathrm{X}_{t-\tau} - \theta_1 \mathrm{X}_{t-\tau}\epsilon_{t-1} \,-\, \cdots \,-\, \theta_q \mathrm{X}_{t-\tau}\epsilon_{t-q}.\end{aligned}$$

Notice that $\mathrm{E}[\epsilon_{t-j}\mathrm{X}_{t-\tau}] = 0$ provided t – j>t – $\tau$, that is, provided j < $\tau$. Write $\mathrm{E}[\epsilon_{t-j}\mathrm{X}_{t-\tau}] = \gamma_{\epsilon x}(\tau - \mathrm{j})$ so that

$$\gamma_{\epsilon x}(\tau) = 0, \ \tau > 0$$

and

$$\gamma_{\epsilon x}(\tau) \neq 0, \ \tau \leq 0.$$

Taking expectations in (11.4.3) we obtain

(11.4.4) $$\gamma_\tau = \phi_1\gamma_{\tau-1}+\cdots+\phi_p\gamma_{\tau-p}+\gamma_{\epsilon x}(\tau) - \theta_1\gamma_{\epsilon x}(\tau-1) - \cdots - \theta_q\gamma_{\epsilon x}(\tau-\text{q}).$$

Hence if $\tau - \text{q} > 0$, that is, $\tau > \text{q}$

(11.4.5) $$\gamma_\tau = \phi_1\gamma_{\tau-1} + \cdots + \phi_p\gamma_{\tau-p}$$

and so

(11.4.6) $$\rho_\tau = \phi_1\rho_{\tau-1} + \cdots + \phi_p\rho_{\tau-p}, \ \tau > \text{q}.$$

We may write this operator form as

(11.4.7) $$\phi(B)\rho_\tau = 0, \ \tau > \text{q}.$$

Thus there are q autocorrelations $\rho_q, \ldots, \rho_1$ which depend on $\theta_1, \ldots, \theta_q$ as well as $\phi_1, \ldots, \phi_p$. The remaining autocorrelations satisfy the usual difference equation, $\phi(B)\rho_\tau = 0$, and, hence, will be an exponentially decaying functions or at least dominated by an exponentially decaying function. When $\tau = 0$,

$$\gamma_0 = \phi_1\gamma_1+\cdots+\phi_p\gamma_p+\sigma_\epsilon^2 - \theta_1\gamma_{\epsilon x}(-1) - \cdots - \theta_q\gamma_{\epsilon x}(-\text{q}).$$

Hence to solve for $\gamma_0$ we must solve simultaneously with the p equations (11.4.4) for $\tau = 1, 2, \ldots, \text{p}$.

Using the covariance generating function the spectral density is

(11.4.8) $$\text{f}(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \frac{|\theta(\text{e}^{-i\lambda})|^2}{|\phi(\text{e}^{-i\lambda})|^2}, \ -\pi \leq \lambda \leq \pi.$$

Because $\theta$(s) and $\phi$(s) are polynomials, this is sometimes referred to as a rational spectral density (rational in polynomials). It is easy to see that rational spectral densities can approximate a very large class of functions, so that the ARMA model is, indeed, a very powerful one. We may expand these polynomials to

(11.4.9) $$\text{f}(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \frac{|1-\theta_1\text{e}^{-i\lambda}-\cdots-\theta_q\text{e}^{-i\lambda q}|^2}{|1-\phi_1\text{e}^{-i\lambda}-\cdots-\phi_p\text{e}^{-i\lambda p}|^2}.$$

Finally, since $\epsilon_t = \theta^{-1}(B)\phi(B)\text{X}_t$ and $\theta^{-1}$(s) is an infinite series, the partial autocorrelation is infinite in extent and eventually behaves like the partial autocorrelation of a pure moving average process. It is perhaps worth examining the ARMA (1, 1) in a bit more detail.

**11.5 The ARMA (1, 1) Process** To understand the generality of the ARMA (p, q) process, it is worthwhile examining the ARMA (1, 1) process. This is particularly useful since the parameter estimation problem for the ARMA (p, q) is in general quite messy and the ARMA (1, 1) illustrates the principles a bit more clearly. The ARMA (1, 1) given by

(11.5.1) $$\text{X}_t - \phi_1\text{X}_{t-1} = \epsilon_t - \theta_1\epsilon_{t-1}.$$

The process is stationary if $-1 < \phi_1 < 1$ and invertible if $-1 < \theta_1 < 1$. The autocovariance is given by

$$\gamma_0 = \phi_1\gamma_1 + \sigma_\epsilon^2 - \theta_1\gamma_{\epsilon x}(-1)$$

(11.5.2)
$$\gamma_1 = \phi_1\gamma_0 - \theta_1\sigma_\epsilon^2$$

$$\gamma_\tau = \phi_1\gamma_{\tau-1}, \quad \tau \geq 2.$$

Also note

$$X_t\epsilon_{t-1} - \phi_1 X_{t-1}\epsilon_{t-1} = \epsilon_t\epsilon_{t-1} - \theta_1\epsilon_{t-1}^2$$

so that taking expectations we obtain

(11.5.3)
$$\gamma_{\epsilon x}(-1) - \phi_1\sigma_\epsilon^2 = -\theta_1\sigma_\epsilon^2$$

Thus, we may substitute the expression in (11.5.3) for $\gamma_{\epsilon z}(-1)$ into (11.5.2) to obtain

(11.5.4)
$$\gamma_0 = \phi_1\gamma_1 + \sigma_\epsilon^2 - \theta_1(\phi_1 - \theta_1)\sigma_\epsilon^2$$

so that using simple algebra

(11.5.5)
$$\gamma_0 - \phi_1\gamma_1 = \sigma_\epsilon^2(1 - \theta_1\phi_1 + \theta_1^2).$$

Also from (11.5.2) we have,

(11.5.6)
$$\phi_1\gamma_1 - \phi_1^2\gamma_0 = \sigma_\epsilon^2(-\theta_1)\phi_1$$

Eliminating the $\phi_1\gamma_1$ term by adding (11.5.5) to (11.5.6),

$$\gamma_0(1 - \phi_1^2) = \sigma_\epsilon^2(1 - \phi_1\theta_1 + \theta_1^2 - \theta_1\phi_1)$$

so that solving for $\gamma_0$ we obtain

(11.5.7)
$$\gamma_0 = \sigma_\epsilon^2 \frac{(1 + \theta_1^2 - 2\theta_1\phi_1)}{1 - \phi_1^2}.$$

By a simple substitution,

(11.5.8)
$$\gamma_1 = \sigma_\epsilon^2 \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2}.$$

By division of these relationships,

$$\rho_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{(1 + \theta_1^2 - 2\phi_1\theta_1)}.$$

and

$$\rho_\tau = \phi_1\rho_{\tau-1}, \ \tau \geq 2.$$

The above discussions allow us to estimate the parameters of an autoregressive process, a moving average process or a mixed process once we have settled on a model. The model building process is not a trivial one and fundamental questions as to model order and type must be addressed. We have suggested the use of the

autocorrelation function and the partial autocorrelation function as tools to the end. The following table summarizes the behavior of different linear models.

| | **Autoregressive** | **Moving Average** | **Mixed (ARMA)** |
|---|---|---|---|
| Models in terms of X's | $\phi(B)X_t = \epsilon_t$ | $\theta^{-1}(B)X_t = \epsilon_t$ | $\theta^{-1}(B)\phi(B)X_t = \epsilon_t$ |
| Models in terms of $\epsilon$'s | $X_t = \phi^{-1}(B)\epsilon_t$ | $X_t = \theta(B)\epsilon_t$ | $X_t = \phi^{-1}(B)\theta(B)\epsilon_t$ |
| $\pi$ Weights | Finite Series | Infinite Series | Infinite Series |
| $\psi$ Weights | Infinite Series | Finite Series | Infinite Series |
| Stationarity Conditions | Roots of $\phi(s) = 0$ lie Outside Unit Circle | Always Stationary | Roots of $\phi(s) = 0$ lie Outside Unit Circle |
| Invertibility Conditions | Always Invertible | Roots of $\theta(s) = 0$ lie Outside Unit Circle | Roots of $\theta(s) = 0$ lie Outside Unit Circle unit circle |
| Autocorrelation Function | Infinite, Tails Off | Finite, Cuts Off | Infinite, Tails Off |
| Partial Autocorrelation Function | Finite, Cuts Off | Infinite, Tails Off | Infinite, Tails Off |

# Chapter 12
# Nonstationary Processes and Time Series

**12.1 Nonstationary Time Series Models** In the previous chapter, we introduced the ARMA model

(12.1.1) $$\phi(B)\mathrm{X}_t = \theta(B)\epsilon_t$$

with $\phi$(s) and $\theta$(s) polynomials in s of degree p and q respectively. We assumed that $\phi$(s) = 0 had roots outside the unit circle so that the resulting process would be stationary. Suppose now we allow the possibility of roots on or inside the unit circle. For simplicity consider the AR(1) model

$$(1 - \phi B)\mathrm{X}_t = \epsilon_t$$

If $\phi$ is taken as two, root is 1/2, then, for example, the series increases or decreases exponentially essentially with no effect from the random input, $\epsilon_t$. We will often refer to the $\epsilon_t$ as *random shocks* or just *shocks.* Similarly in any model for which one or more of the roots of $\phi$(s) = 0 lie strictly within the unit circle, the process will behave in this exponential manner essentially independent of the $\epsilon_t$'s. In this circumstance the process is essentially deterministic and there is little need for a statistical analysis. We consider then only those cases where the roots lie on the unit circle.

In particular, we now entertain the model

(12.1.2) $$\phi(B)(I - B)^d\mathrm{X}_t = \theta(B)\epsilon_t.$$

Here $\phi(B)$ is an ordinary stationary autoregressive operator. We write $I - B = \Delta$, the difference operator. Thus our model can be written $\phi(B)\Delta^d\mathrm{X}_t = \theta(B)\epsilon_t$ or $\phi(B)\mathrm{Y}_t = \theta(B)\epsilon_t$ with $\mathrm{Y}_t = \Delta^d\mathrm{X}_t$. After proper differencing of the process $\mathrm{X}_t$ we have the usual ARMA (p, q) model. Now $\mathrm{X}_t = \mathrm{S}^d\mathrm{Y}_t$ where S is the infinite summation operator defined by

(12.1.3) $$\mathrm{SX}_t = \sum_{k=-\infty}^{t} \mathrm{X}_k = (1+B+B^2+B^3+\cdots)\,\mathrm{X}_t$$

which may be written

(12.1.4) $$\mathrm{SX}_t = (1 - B)^{-1}\mathrm{X}_t = \Delta^{-1}\mathrm{X}_t$$

Thus the summation operator is the inverse of the difference operator, $\Delta$. Similarly

$$\mathrm{S}^2\mathrm{X}_t = \mathrm{SX}_t + \mathrm{SX}_{t-1} + \mathrm{SX}_{t-2}+\cdots$$

which may be written as

$$\mathrm{S}^2 = \sum_{i=-\infty}^{t}\sum_{h=-\infty}^{i} \mathrm{X}_h$$

and

$$\mathrm{S}^3\mathrm{X}_t = \sum_{j=-\infty}^{t}\sum_{i=-\infty}^{j}\sum_{h=-\infty}^{i} \mathrm{X}_h.$$

Differencing is very much analogous to differentiation while summing is analogous to integration. For this reason the model (12.1.2) is called the *autoregressive integrated moving average process of order (p, d, q)* and abbreviated ARIMA(p, d, q).

One important feature of the nonstationary process of the type we seek to analyze with the ARIMA model is that the local behavior must be independent of the level of the process. That is, we want,

(12.1.5) $$\phi(B)(X_t+c) = \phi(B)X_t.$$

Note that by adding the $\Delta$ operator in (12.1.5), we obtain

$$\phi(B)\Delta(X_t+c) = \phi(B)[X_t+c - X_{t-1} - c] = \phi(B)\Delta X_t$$

which is independent of level. Thus, for a process which is homogeneous except in level, the model

(12.1.6) $$\phi(B)\Delta X_t = \theta(B)\epsilon_t$$

is appropriate. If, in contrast, a series has neither fixed level nor fixed slope, we require

$$\phi(B)(X_t+mt+c) = \phi(B)X_t$$

By introducing $\Delta^2$ into this equation, we have

$$\phi(B)\Delta^2(X_t+mt+c) = \phi(B)\Delta^2 X_t.$$

In this case an ARIMA (p, 2, q) is appropriate.

It is sometimes useful to consider a slight extension of the ARIMA process

(12.1.7) $$\phi(B)\Delta^d X_t = \theta_0+\theta(B)\epsilon_t$$

The operator $\phi(B)$ is the autoregressive operator and it is assumed that the roots lie outside the unit circle. The operator $\phi(B)\Delta^d$ is the *generalized autoregressive operator* and finally $\theta(B)$ is the *moving average operator* and it is assumed that the roots here also lie outside the unit circle. If we wish to include a *deterministic* function of time as a trend, this can be done by choosing $\theta_0$ non-zero. For example if d = 1, we can estimate a deterministic linear trend in the presence of nonstationary noise.

**12.2 Some Important Special Cases** We consider the following special cases in some detail: *i.* ARIMA (0, 1, 1), $\Delta X_t = \epsilon_t - \theta_1\epsilon_{t-1} = (I - \theta_1 B)\epsilon_t$; *ii.* ARIMA (0, 2, 2), $\Delta^2 X_t = a_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} = (I - \theta_1 B - \theta_2 B^2)\epsilon_t$; and *iii.* ARIMA (1, 1, 1), $\Delta X_t - \phi_1\Delta X_{t-1} = \theta_1\epsilon_{t-1}$ or $(I - \phi_1 B)\Delta X_t = (I - \theta_1 B)\epsilon_t$. Let us now write for the generalized autoregressive operator,

(12.2.1) $$\Phi(B) = \phi(B)\Delta^d.$$

We will write the model in three forms.

Let $\Phi(B) = I - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_{p+d} B^{p+d}$, so that the general model ($\theta_0 = 0$) can be written

(12.2.2) $$X_t = \Phi_1 X_{t-1} + \cdots + \Phi_{p+d} X_{t-p-d} - \theta_1\epsilon_{t-1} - \cdots - \theta_q\epsilon_{t-q} + \epsilon_t$$

For example, the ARIMA (1, 1, 1) is $(1 - \phi B)(1 - B)X_t = (1 - \theta B)\epsilon_t$ where we drop the suffix 1 on $\phi_1$ and $\theta_1$. This ARIMA (1, 1, 1) model may be expanded as

$$\{I - (1+\phi)B+\phi B^2\}X_t = (I - \theta B)\epsilon_t$$

or

$$X_t = (1+\phi)X_{t-1} - \phi X_{t-2}+\epsilon_t - \theta\epsilon_{t-1}.$$

It is sometimes convenient to represent the ARIMA models as a infinite moving average, that is, with $\psi$ weights. For nonstationary time series models, the $\psi$-weight representation does not converge. Thus strictly

speaking the following discussion is only a formal computation. It may be used to suggest relationships among the $\pi$-weights, the $\psi$-weights, the $\theta$-weights and the $\phi$-weights, but because we are manipulating non-convergent series, these relationships must be verified by more rigorous calculations. We may formally write $X_t$ in the infinite moving average form

$$X_t = \epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2} + \cdots$$

so that

$$X_t = \psi(B)\epsilon_t.$$

Operating with $\Phi(B)$ on both sides yields

$$\Phi(B)\, X_t = \Phi(B)\psi(B)\epsilon_t$$

But since

$$\Phi(B)\, X_t = \theta(B)\epsilon_t$$

we can identify $\theta(B) = \Phi(B)\psi(B)$. Thus,

$$(I - \Phi_1 B - \cdots - \Phi_{p+d}B^{p+d})(I + \psi_1 B + \psi_2 B^2 + \cdots) = (I - \theta_1 B - \cdots - \theta_q B^q)$$

Notice that if j is sufficiently large, namely j $\geq$ max(p+d $-$ 1, q), then the $\psi$ weights must satisfy

$$\Phi(B)\psi_j = 0.$$

Thus for j sufficiently large, the $\psi_j$ satisfy the same homogeneous difference equation as the covariance function, $\gamma_\tau$, and, hence, can be written as a mixture of polynomials, exponentials and sinusoids in the argument j.

Consider for example the ARIMA (1, 1, 1). Here

$$\Phi(B) = I - (1+\phi)B + \phi B^2$$

so that

$$\{I - (1+\phi)B + \phi B^2\}(1 + \psi_1 B + \psi_2 B^2 + \cdots) = I - \theta B.$$

Thus

$$-(1+\phi) + \psi_1 = -\theta$$

and for $\psi_j \geq 1$,

$$\psi_j - (1+\phi)\psi_{j-1} + \theta\psi_{j-2} = 0.$$

This is a second order difference equation with roots 1 and $\phi^{-1}$. Thus $\psi_j = A_0 + A_1\phi^j$, j $\geq$ 1. Solving this simultaneously with $-(1+\phi)+\psi_1 = -\theta$ yields

$$A_0 = \tfrac{1-\theta}{1-\phi} \text{ and } A_1 = \tfrac{\theta-\phi}{1-\phi}.$$

Thus the ARIMA (1, 1, 1) can be written

$$X_t = \sum_{j=0}^{\infty}(A_0 + A_1\phi^j)\epsilon_{t-j}$$

with the above values for $A_0$ and $A_1$. Notice that for $A_0 \neq 0$, this is not a convergent series and so, in some sense this infinite series representation is nonsense. Even so, it will have some utility in establishing forecasting procedures for the nonstationary time series models.

Finally we can write the ARIMA time series model in its inverted form $\pi(B)X_t = \epsilon_t$ or

$$X_t = \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots + \epsilon_t.$$

As before, operating on both sides with $\theta(B)$

$$\theta(B)\pi(B)X_t \;=\; \theta(B)\epsilon_t \;=\; \Phi(B)X_t.$$

Thus we have $\theta(B)\pi(B) \;=\; \Phi(B)$ or

$$(I - \Phi_1 B - \cdots - \Phi_{p+d}B^{p+d}) \;=\; (I - \theta_1 B - \cdots - \theta_q B)(I - \pi_1 B - \pi_2 B^2 - \cdots).$$

For j > max(p+d, q), $\theta(B)\pi_j = 0$. We again return to our ARIMA (1, 1, 1) as an example. We have

(12.2.3) $$\theta(B)\pi(B) \;=\; \Phi(B)$$

or

$$\pi(B) \;=\; \{I - (1+\phi)B+\phi B^2\}(I+\theta B+\theta^2 B^2+\cdots)$$

so that by equating corresponding coefficients

$$-\pi_1 \;=\; \theta - 1 - \phi$$
$$-\pi_2 \;=\; \phi+\theta^2 - \theta(1+\phi)$$

and

$$-\pi_j \;=\; \theta^j - (1+\phi)\theta^{j-1}+\phi\theta^{j-2}, \quad j \;\geq\; 3.$$

Rewriting and solving for the $\pi$-weights

$$\pi_1 \;=\; 1+\phi - \theta$$
$$\pi_2 \;=\; (\theta - \phi)(1 - \theta)$$
$$\pi_j \;=\; (\theta - \phi)(1 - \theta)\theta^{j-2}, \; j \;\geq\; 3.$$

Let us also observe that if d is greater than 0, since $\Phi(s) = \phi(s)(1 - s)^d$, we may conclude that $\Phi(1) \;=\; 0$. Substituting in (12.2.3)

$$0 \;=\; \Phi(1) \;=\; \theta(1)\pi(1).$$

Since $\theta(s) \;=\; 0$ has roots outside the unit circle, $\theta(1) \;\neq\; 0$. Thus $\pi(1) \;=\; 0$ so that $1 - \sum_{j=1}^{\infty}\pi_j \; 1^j \;=\; 0$ or equivalently

(12.2.4) $$\sum_{j=1}^{\infty} \pi_j \;=\; 1$$

Thus if d $>$ 0, the $\pi$-weights must sum to 1. They are not necessarily positive, so that they may not be a probability distribution.

**12.3 Integrated Moving Average Process Models** In this section we wish to consider *integrated moving average* processes (IMA) of the form $\Delta^d X_t \;=\; \theta(B)\epsilon_t$. Two models of particular interest are the ARIMA (0, 1, 1) $=$ IMA (1, 1), $X_t \;=\; X_{t-1}+\epsilon_t - \theta\epsilon_{t-1}$ and the ARIMA(0, 2, 2) $=$ IMA (2, 2), $X_t \;=\; 2X_{t-1} - X_{t-2}+\epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2}$.

First the IMA(1, 1). Let us note that we may write

$$I - \theta B \;=\; (1 - \theta)B+(I - B) \;=\; (1 - \theta)B+\Delta \;=\; \eta B+\Delta, \text{ with } \eta \;=\; 1 - \theta.$$

We can thus reparametrize the process as

$$\Delta X_t \;=\; \eta\epsilon_{t-1}+\Delta\epsilon_t$$

Again the following computation is purely formal. Applying $\Delta^{-1} \;=\; S$ to both sides

$$X_t \;=\; \eta S\epsilon_{t-1}+\epsilon_t$$

But $\Delta^{-1} \;=\; (1 - B)^{-1} \;=\; 1+B+B^2+\cdots$ so that

$$X_t \;=\; \eta\sum_{j=1}^{\infty}\epsilon_{t-j} + \epsilon_t.$$

The $\psi$-weights are $\psi_0 \;=\; 1$ and $\psi_j \;=\; \eta \;=\; 1 - \theta$, j $\geq$ 1. To consider the inverted form of the model

$$X_t \;=\; \sum_{j=1}^{\infty}\pi_j X_{t-j}+\epsilon_t \;=\; \bar{X}_{t-1}(\pi) + \epsilon_t$$

where we define $\bar{X}_{t-1}(\pi)$ to be $\sum_{j=1}^{\infty}\pi_j X_{t-j}$. We notice

$$(I - \theta B)\pi(B) \;=\; I \;-\; B.$$

Equivalently,

$$\pi(\mathrm{s}) \;=\; \tfrac{1-\mathrm{s}}{1-\theta\mathrm{s}} \;=\; \tfrac{1-\theta\mathrm{s}-(1-\theta)\mathrm{s}}{1-\theta\mathrm{s}}.$$

We may simplify this to obtain

$$\pi(\mathrm{s}) \;=\; 1 - \tfrac{(1-\theta)\mathrm{s}}{1-\theta\mathrm{s}}.$$

Expanding the denominator as an infinite series

$$\pi(\mathrm{s}) \;=\; 1 \;-\; \mathrm{s}(1-\theta)(1+\theta\mathrm{s}+\theta^2\mathrm{s}^2+\cdots).$$

or

$$\pi(\mathrm{s}) \;=\; 1 \;-\; (1-\theta)(\mathrm{s}+\theta\mathrm{s}^2+\theta^2\mathrm{s}^3+\cdots).$$

But $\pi_j$ is the negative of the coefficient of $\mathrm{s}^j$ so that

$$\pi_j \;=\; (1-\theta)\theta^{j-1} \;=\; \eta(1-\eta)^{j-1},\; \mathrm{j} \;\geq\; 1$$

except for $\pi_0$ which is 1. Thus $\mathrm{X}_t$ is an *exponentially weighted moving average* (EWMA) of previous values. This is sometimes also called an *exponential smoother* or an *exponential filter*. For this case, we can write

$$\overline{\mathrm{X}}_{t-1}(\eta) \;=\; \eta\sum_{j=1}^{\infty}(1-\eta)^{j-1}\mathrm{X}_{t-j}$$

so that

$$\overline{\mathrm{X}}_{t}(\eta) \;=\; \eta\sum_{j=1}^{\infty}(1-\eta)^{j-1}\mathrm{X}_{t+1-j}.$$

Factoring out the first term,

$$\overline{\mathrm{X}}_{t}(\eta) \;=\; \eta\mathrm{X}_t + \eta\sum_{j=2}^{\infty}(1-\eta)^{j-1}\mathrm{X}_{t+1-j}.$$

We can make a change of index by letting $\mathrm{k} = \mathrm{j} - 1$ so that

$$\overline{\mathrm{X}}_{t}(\eta) \;=\; \eta\mathrm{X}_t + \eta\sum_{k=1}^{\infty}(1-\eta)^{k}\,\mathrm{X}_{t-k}.$$

Factoring out $(1-\eta)$ we have

$$\overline{\mathrm{X}}_{t}(\eta) \;=\; \eta\mathrm{X}_t + (1-\eta)\eta\sum_{k=1}^{\infty}(1-\eta)^{k-1}\mathrm{X}_{t-k}.$$

Thus

$$\overline{\mathrm{X}}_{t}(\eta) \;=\; \eta\mathrm{X}_t + (1-\eta)\,\overline{\mathrm{X}}_{t-1}(\eta).$$

The function, $\overline{\mathrm{X}}_{t}(\eta)$, may be thought of as the "level" of the process at time t. Since $-1 < \theta < 1$, we have that $0 < \eta < 2$. In most cases $0 < \eta < 1$. If $\eta$ is close to 0, the "level" is almost constant. If $\eta$ is close to 1, the level is determined almost exclusively by the most recent observation, $\mathrm{X}_t$. Since

$$\mathrm{X}_t \;=\; \overline{\mathrm{X}}_{t-1}(\eta) + \epsilon_t$$

and

$$\overline{\mathrm{X}}_{t}(\eta) \;=\; \eta\mathrm{X}_t + (1-\eta)\,\overline{\mathrm{X}}_{t-1}(\eta)$$

we have

$$\overline{\mathrm{X}}\,(\eta) \;=\; \eta\overline{\mathrm{X}}_{t-1}(\eta) + \eta\epsilon_t + (1-\eta)\,\overline{\mathrm{X}}_{t-1}(\eta)$$

or

$$\bar{X}_t(\eta) = \bar{X}_{t-1}(\eta) + \eta\epsilon_t.$$

Thus while the process itself has the random shock, $\epsilon_t$, the level of the process has random shock, $\eta\epsilon_t$, which may be more or less than $\epsilon_t$ depending on $\eta$.

Let us now turn to the ARIMA (0, 2, 2). We write

$$I - \theta_1 B - \theta_2 B^2 = (\eta_0\Delta+\eta_1)B+\Delta^2$$

where

$$\theta_1 = 2 - \eta_0 - \eta_1 \text{ and } \theta_2 = \eta_0 - 1$$

or in inverse form

$$\eta_0 = 1+\theta_2 \text{ and } \eta_1 = 1 - \theta_1 - \theta_2.$$

We can then write the ARIMA (0, 2, 2) as

$$\Delta^2 X_t = (\eta_0\Delta+\eta_1)\epsilon_{t-1} + \Delta^2\epsilon_t.$$

Again we make a formal computation. Taking inverse $\Delta$, $\Delta^{-1} = S$, twice

$$X_t = \eta_0 S\epsilon_{t-1} + \eta_1 S^2\epsilon_{t-1} + \epsilon_t.$$

But

$$S\epsilon_{t-1} = \epsilon_{t-1} + \epsilon_{t-2} + \epsilon_{t-3} + \cdots$$

so that

$$S^2\epsilon_{t-1} = S\epsilon_{t-1} + S\epsilon_{t-2} + S\epsilon_{t-3} + \cdots.$$

We may expand to get

$$\begin{aligned} S^2\epsilon_{t-1} = {} & \epsilon_{t-1} + \epsilon_{t-2} + \epsilon_{t-3} + \cdots \\ & + \epsilon_{t-2} + \epsilon_{t-3} + \cdots \\ & + \epsilon_{t-3} + \epsilon_{t-4} + \cdots \\ & + \epsilon_{t-4} + \epsilon_{t-5} + \cdots \\ & \vdots \end{aligned}$$

Collecting like terms, we have

$$S^2\epsilon_{t-1} = \epsilon_{t-1} + 2\epsilon_{t-2} + 3\epsilon_{t-3} + \cdots.$$

Thus

$$X_t = \epsilon_t + (\eta_0+1\eta_1)\epsilon_{t-1} + (\eta_0+2\eta_1)\epsilon_{t-2} + (\eta_0+ 3\eta_1)\epsilon_{t-3} + \cdots.$$

The $\psi$ weights are

$$\begin{aligned} \psi_0 &= 1 \\ \psi_j &= \eta_0 + j\eta_1, \ j \geq 1 \end{aligned}$$

To find the inverted form, we equate coefficients in

$$(1 - 2s+s^2) = (1 - \theta_1 s - \theta_2 s^2)(1 - \pi_1 s - \pi_2 s^2 - \cdots).$$

Thus

$$\begin{aligned} \pi_1 &= 2 - \theta_1 = \eta_0+\eta_1 \\ \pi_2 &= \theta_1(2 - \theta_1) - (1+\theta_2) = \eta_0 + 2\eta_1 - (\eta_0+\eta_1)^2 \end{aligned}$$

and

$$(1 - \theta_1 B - \theta_2 B^2)\pi_j = 0, \quad j \geq 3.$$

# Chapter 13
# Least Squares Forecasts

**13.1 Linear Forecasting Functions.** We now wish to consider forecasting based on the general ARIMA process model

$$\Phi(B)X_t = \phi(B)\Delta^d X_t = \theta(B)\epsilon_t.$$

We suppose we are currently at time t and we wish to forecast m steps ahead, $X_{t+m}$. We shall write the forecast of $X_{t+m}$ as $\hat{X}_t(m)$. We wish to write $\hat{X}_t(m)$ as a linear function of the past values of the time series data we know, that is, $X_t, X_{t-1}, X_{t-2}, \ldots$ . Equivalently we may write $\hat{X}_t(m)$ as a linear function of $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots$ . We will assume an infinite moving average representation exists. For a nonstationary model this representation doesn't converge. Hence, for nonstationary models, the following discussion is a formal representation, but not a rigorous proof.

We wish to make a linear representation of the process depending on the past values $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots$ which we do by setting

$$\hat{X}_t(m) = \psi^*_m\epsilon_t + \psi^*_{m+1}\epsilon_{t-1} + \psi^*_{m+2}\epsilon_{t-2} + \cdots. \tag{13.1.1}$$

Here, the $\psi^*_j$ are weights on the $\epsilon_t$ which may or may not be related to the $\psi_j$ weights. We do know however that

$$X_{t+m} = \epsilon_{t+m} + \psi_1\epsilon_{t+m-1} + \cdots + \psi_m\epsilon_t + \psi_{m+1}\epsilon_{t-1} + \cdots. \tag{13.1.2}$$

Subtracting $\hat{X}_t(m)$ from $X_{t+m}$, squaring and then taking expectations yields

$$E[X_{t+m} - \hat{X}_t(m)]^2 = (1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{m-1}^2)\sigma_\epsilon^2 + \sum_{j=0}^{\infty}\left\{\psi_{m+j} - \psi^*_{m+j}\right\}^2\sigma_\epsilon^2.$$

This may be minimized by choosing $\psi^*_{m+j} = \psi_{m+j}$. Thus the forecast is

$$\hat{X}_t(m) = \psi_m\epsilon_t + \psi_{m+1}\epsilon_{t-1} + \psi_{m+2}\epsilon_{t-2} + \cdots. \tag{13.1.3}$$

We then may write

$$e_t(m) = \epsilon_{t+m} + \psi_1\epsilon_{t+m-1} + \cdots + \psi_{m-1}\epsilon_{t+1}$$

so that

$$X_{t+m} = e_t(m) + \hat{X}_t(m). \tag{13.1.4}$$

We call $e_t(m)$ the *forecast error* of lead time m. Let us write $E[X_{t+m} \mid X_t, X_{t-1}, \cdots]$ as $E[X_{t+m} \mid t]$. We first notice that

$$E[X_{t+m} \mid t] = E[\sum_{j=0}^{\infty}\psi_j\epsilon_{t+m-j} \mid t] = \sum_{j=0}^{\infty}\psi_j\, E[\epsilon_{t+m-j} \mid t] = \sum_{j=m}^{\infty}\psi_j\epsilon_{t+m-j} = \hat{X}_t(m).$$

Hence $\hat{X}_t(m)$ is the best linear estimator of $X_{t+m}$ given the data up to time, t. We may now examine

$$E[e_t(m) \mid t] = E[\epsilon_{t+m} + \psi_1\epsilon_{t+m-1} + \cdots + \psi_{m-1}\epsilon_{t+1} \mid t] = 0. \tag{13.1.5}$$

Hence the expected forecast error is 0 and $\hat{X}_t(m)$ is the minimum mean square error estimator and is unbiased. Hence $\hat{X}_t(m)$ is the *best* estimator in the Rao-Blackwell sense. Let us also notice that

$$\text{var}[e_t(m) \mid t] \;=\; (1 + \psi_1^2 + \cdots + \psi_{m-1}^2)\sigma_\epsilon^2.$$

**Forecasting Linear Combinations of $X_t$**

Consider now the array, $X_{t+1},\ldots,X_{t+m}$ defined as follows

$$\begin{aligned} X_{t+1} &= \epsilon_{t+1} + \psi_1\epsilon_t + \psi_2\epsilon_{t-1} + \cdots \\ &\vdots \\ X_{t+m} &= \epsilon_{t+m} + \psi_1\epsilon_{t+m-1} + \psi_m\epsilon_t + \psi_{m+1}\epsilon_{t-1} + \cdots . \end{aligned}$$

We may define Q by

$$Q \;=\; \sum_{j=1}^{m} w_j X_{t+j} \;=\; w_1 X_{t+1} + w_2 X_{t+2} + \cdots + w_m X_{t+m}.$$

Writing the $X_t$ in terms of their $\psi$-weight representations we have

$$\begin{aligned} Q = {} & w_m\epsilon_{t+m} + (w_{m-1} + \psi_1 w_m)\epsilon_{t+m-1} + \cdots + (w_1 + \psi_1 w_2 + \cdots + \psi_{m-1}w_m)\epsilon_{t+1} \\ & + (\psi_1 w_1 + \cdots + \psi_m w_m)\epsilon_t \;+\; (\psi_2 w_2 + \cdots + \psi_{m+1} w_m)\epsilon_{t-1} + \cdots . \end{aligned}$$

Suppose we want a linear forecast of Q in the form $Q \;=\; c_1\epsilon_t + c_2\epsilon_{t-1} + \cdots$. Then as before

$$\begin{aligned} E[Q - \hat{Q}]^2 \;=\; & [w_m^2 + (w_{m-1} + \psi_1 w_m)^2 + \cdots + (w_1 + \psi_1 w_2 + \cdots + \psi_{m-1}w_m)^2]\sigma_\epsilon^2 \\ & + \sigma_\epsilon^2\Big\{(\psi_1 w_1 + \cdots + \psi_m w_m)^2 \;+\; (\psi_2 w_2 + \cdots + \psi_{m+1} w_m)^2 + \cdots\Big\} . \end{aligned}$$

This may be minimized by choosing $c_1 = (\psi_1 w_1 + \cdots + \psi_m w_m)$, $c_2 \;=\; (\psi_2 w_2 + \cdots + \psi_{m+1} w_m)$, and so on. Thus the minimum mean square error forecast is

$$\hat{Q} = (\psi_1 w_1 + \cdots + \psi_m w_m)\epsilon_t + (\psi_2 w_2 + \cdots + \psi_{m+1} w_m)\epsilon_{t-1} + \cdots.$$

But from our earlier discussion, we know

$$\begin{aligned} \hat{X}_t(1) &= \psi_1\epsilon_t + \psi_2\epsilon_{t-1} + \cdots \\ \hat{X}_t(2) &= \psi_2\epsilon_t + \psi_3\epsilon_{t-1} + \cdots \\ &\vdots \\ \hat{X}_t(m) &= \psi_m\epsilon_t + \psi_{m+1}\epsilon_{t-1} + \cdots. \end{aligned}$$

From this we may conclude that

$$\hat{Q} \;=\; w_1\hat{X}_t(1) \;+\; \cdots \;+\; w_m\hat{X}_t(m). \tag{13.1.6}$$

Thus the least squares forecast of $w_1 X_{t+1} + \;\cdots\; + w_m X_{t+m}$ is $w_1\hat{X}_t(1) \;+\; \cdots \;+\; w_m\hat{X}_t(m)$.

**Correlations of Forecast Errors.** Now let us consider the correlations of forecast errors. First we note that

$$e_t(1) \;=\; X_{t+1} - \hat{X}_t(1) \;=\; \psi_0\epsilon_{t+1} \;=\; \epsilon_{t+1}.$$

The one-step-ahead forecast errors are uncorrelated. Also we note

$$\mathrm{e}_t(\mathrm{m}) = \mathrm{X}_{t+m} - \hat{\mathrm{X}}_t(\mathrm{m}) = \epsilon_{t+m} + \psi_1\epsilon_{t+m-1} + \cdots + \psi_{m-1}\epsilon_{t+1}$$

and

$$\mathrm{e}_{t-j}(\mathrm{m}) = \mathrm{X}_{t+m-j} - \hat{\mathrm{X}}_{t-j}(\mathrm{m}) = \epsilon_{t+m-j} + \psi_1\epsilon_{t+m-j-1} + \cdots + \psi_{m-1}\epsilon_{t-j+1}.$$

Thus

$$\mathrm{E}[\mathrm{e}_t(\mathrm{m})\mathrm{e}_{t-j}(\mathrm{m})] = [\psi_0\psi_j + \psi_1\psi_{j+1} + \cdots + \psi_{m-1-j}\psi_{m-1}]\sigma_\epsilon^2$$

$$= \sigma_\epsilon^2 \sum_{i=j}^{m-1} \psi_i\psi_{i-j} \text{ with } \psi_0 = 1.$$

Thus

$$\mathrm{corr}\Big\{\mathrm{e}_t(\mathrm{m}),\, \mathrm{e}_{t-j}(\mathrm{m})\Big\} = \frac{\sum_{i=j}^{m-1}\psi_i\psi_{i-j}}{\sum_{i=0}^{m-1}\psi_i^2},\ 0 \le \mathrm{j} < \mathrm{m}$$

and, of course, 0 elsewhere.

Finally, we observe

$$\begin{aligned}\mathrm{e}_t(\mathrm{m}+\mathrm{j}) &= \mathrm{X}_{t+m-j} - \hat{\mathrm{X}}_t(\mathrm{m}+\mathrm{j}) \\ &= \epsilon_{t+m+j} + \psi_1\epsilon_{t+m+j-1} + \cdots + \psi_j\epsilon_{t+m} + \psi_{j+1}\epsilon_{t+m-1} + \cdots + \psi_{m+j-1}\epsilon_{t+1}.\end{aligned}$$

Thus

$$\mathrm{E}[\mathrm{e}_t(\mathrm{m})\mathrm{e}_t(\mathrm{m}+\mathrm{j})] = [\psi_0\psi_j + \psi_1\psi_{j+1} + \cdots + \psi_{m-1}\psi_{m+j-1}]\sigma_\epsilon^2$$

so that

$$\mathrm{E}[\mathrm{e}_t(\mathrm{m})\mathrm{e}_t(\mathrm{m}+\mathrm{j})] = \sigma_\epsilon^2 \sum_{i=0}^{m-1} \psi_i\psi_{i+j}.$$

From this we may compute the correlation,

$$\mathrm{corr}\Big\{\mathrm{e}_t(\mathrm{m}),\, \mathrm{e}_t(\mathrm{m}+\mathrm{j})\Big\} = \frac{\sum_{i=0}^{m-1}\psi_i\psi_{i+j}}{\left\{\sum_{k=0}^{m-1}\psi_k^2 \sum_{i=0}^{m-j-1}\psi_i^2\right\}^{\frac{1}{2}}}.$$

**13.2 Other Forms of Forecast Functions.** We have so far considered forecasts from the $\psi$-weight formulation of the general linear model. Let us now consider some alternatives. We first observe that

$$\begin{aligned}&\mathrm{E}[\mathrm{X}_{t-j} \mid \mathrm{t}] = \mathrm{X}_{t-j},\ \mathrm{j} = 0, 1, 2, \ldots \\ &\mathrm{E}[\mathrm{X}_{t+j} \mid \mathrm{t}] = \hat{\mathrm{X}}_t(\mathrm{j}),\ \mathrm{j} = 1, 2, 3, \ldots \\ &\mathrm{E}[\epsilon_{t-j} \mid \mathrm{t}] = \epsilon_{t-j} = \mathrm{X}_{t-j} - \hat{\mathrm{X}}_{t-j-1}(1),\ \mathrm{j} = 0, 1, 2, \ldots\end{aligned}$$

and finally

$$E[\epsilon_{t+j} \mid t] \;=\; 0,\; j \;=\; 1, 2, 3, \ldots\ .$$

Consider the mixed formulation

$$X_{t+m} = \Phi_1 X_{t+m-1} + \cdots + \Phi_{p+d} X_{t+m-p-d} - \theta_1 \epsilon_{t+m-1} - \cdots - \theta_q \epsilon_{t+m-q} + \epsilon_{t+m}.$$

Thus we have

(13.2.1)
$$\begin{aligned}\hat{X}_t(m) = {} & \Phi_1 E[X_{t+m-1} \mid t] + \cdots + \Phi_{p+d} E[X_{t+m-p-d} \mid t] \\ & - \theta_1 E[\epsilon_{t+m-1} \mid t] - \cdots - \theta_q E[\epsilon_{t+m-q} \mid t] + E[\epsilon_{t+m} \mid t].\end{aligned}$$

Using the above forms of the conditional expectations, we can use (13.2.1) as a forecast function. To make this clearer, it is worth considering some examples. Consider for example $(I - .8\,B)(I - B)X_{t+m} = \epsilon_{t+m}$. Then

$$X_{t+m} = \; 1.8\; X_{t+m-1} \; - \; .8\; X_{t+m-2} \; + \epsilon_{t+m}.$$

Then

$$\begin{aligned}\hat{X}_t(1) \;&=\; 1.8\; X_t \;-\; .8\; X_{t-1} \\ \hat{X}_t(2) \;&=\; 1.8\; \hat{X}_t(1) \;-\; .8\; X_t\end{aligned}$$

and

$$\hat{X}_t(m) \;=\; 1.8\; \hat{X}_t(m-1) \;-\; .8\; \hat{X}_t(m-2), \quad m \;=\; 3, 4, 5, \ldots\ .$$

As another example, consider

$$\Delta^2 X_{t+m} \;=\; (I \;-\; .9\; B \;-\; .5\; B^2)\; \epsilon_{t+m}.$$

Then

$$X_{t+m} = 2\; X_{t+m-1} \;-\; X_{t+m-2} \;+\; \epsilon_{t+m} \;-\; .9\; \epsilon_{t+m-1} \;+\; .5\; \epsilon_{t+m-2}.$$

We may thus apply the conditional expectations to obtain

$$\begin{aligned}\hat{X}_t(1) &= 2\; X_t - X_{t-1} - .9\; [X_t - \hat{X}_{t-1}(1)] + .5\; [X_{t-1} - \hat{X}_{t-2}(1)] \\ \hat{X}_t(2) &= 2\; \hat{X}_t(1) - X_t - .9\; [X_t - \hat{X}_{t-1}(1)]\end{aligned}$$

and

$$\hat{X}_t(m) = 2\; \hat{X}_t(m-1) - \hat{X}_t(m-2).$$

Notice that in both cases, eventually, $\Phi(B)\hat{X}_t(m) = 0$ where the operator B operates on the index, m. Thus eventually the forecast $\hat{X}_t(m)$ has the same structure as the covariance, $\gamma_\tau$. Finally we note that we can write

$$X_{t+m} = \sum_{j=1}^{\infty} \pi_j X_{t+m-j} \;+\; \epsilon_{t+m}$$

so that

(13.2.2)
$$\hat{X}_t(m) \;=\; \sum_{j=1}^{m-1} \pi_j \hat{X}_t(m-j) \;+\; \sum_{j=m}^{\infty} \pi_j X_{t+m-j}$$

**13.3 Practical Forecasting** We have now developed three forms of the forecasting function, specifically

(13.3.1)
$$\hat{X}_t(m) = \psi_m E[\epsilon_t \mid t] + \psi_{m+1} E[\epsilon_{t-1} \mid t] + \cdots,$$

(13.3.2)
$$\begin{aligned}\hat{X}_t(m) = {} & \Phi_1 E[X_{t+m-1} \mid t] + \cdots + \Phi_{p+d} E[X_{t+m-p-d} \mid t] \\ & - \theta_1 E[\epsilon_{t+m-1} \mid t] - \cdots - \theta_q E[\epsilon_{t+m-q} \mid t] + E[\epsilon_{t+m} \mid t],\end{aligned}$$

and

$$\hat{X}_t(m) = \sum_{j=1}^{m-1} \pi_j \hat{X}_t(m-j) + \sum_{j=m}^{\infty} \pi_j X_{t+m-j}. \tag{13.3.3}$$

Equation (13.3.2) is a useful computational algorithm if the process is purely autoregressive in form. However, if there is a nontrivial moving average component, the computation of $E[\epsilon_{t-j} \mid t]$ involves $X_{t-j} - \hat{X}_{t-j-1}(1)$. This leads to an infinite regression and, so, creates a problem with initializing the algorithm. The same is basically true of (13.3.1). Equation (13.3.3) can be used, but of course we would generally only have data $X_1, \ldots, X_t$ so that

$$\hat{X}_t(m) = \sum_{j=1}^{m-1} \pi_j \hat{X}_t(m-j) + \sum_{j=m}^{t+m-1} \pi_j X_{t+m-j}$$

is the practical formula. Because of the truncation of the second series, this would only be practical in settings where the $\pi$-weights converged to 0 sufficiently rapidly that the truncated part of the series were negligible. Clearly the model for the time series must be invertible for this to be true, but even more the time series must be sufficiently long so that we can reach sufficiently far back into the tail of the second series.

To calculate the $\pi$-weights, we use the standard device we have used earlier. Since $\Phi(B)X_t = \theta(B)\epsilon_t$ and also $\Pi(B)X_t = \epsilon_t$, we can substitute the second in the first to obtain

$$\Phi(B)X_t = \theta(B)\Pi(B)X_t.$$

By equating operators, we get $\Phi(B) = \theta(B)\Pi(B)$. We can then equate the coefficients of the Bs to solve for $\pi_j$. In general

$$\begin{aligned}
&\pi_1 = \Phi_1 - \theta_1 \\
&\pi_2 = \Phi_2 - \theta_2 + \theta_1(\Phi_1 - \theta_1) \\
&\vdots \\
&\theta(B)\pi_j = 0,\ j > \max\{p+d,\ q\}.
\end{aligned}$$

In the last difference equation, the operator $\theta(B)$ operates on the j subscript. Notice that if the roots of the auxiliary equation, $\theta(s) = 0$, lie outside the unit circle, then the $\pi$-weights will be exponentially damped (or dominated by something which is exponentially damped) and, hence, we will be able to use the $\pi$-weight representation to initiate our forecasting.

**13.4 Updating Forecasts** One issue of substantial interest is an updating procedure. That is, if we know the time series up to time, t, e.g. $X_1, \ldots, X_t$, and we add a new observation, $X_{t+1}$, it is desirable not to have to recompute the whole forecast from scratch. To see how to update, recall that

$$\hat{X}_{t+1}(m) = \psi_m \epsilon_{t+1} + \psi_{m+1}\epsilon_t + \psi_{m+2}\epsilon_{t-1} + \cdots$$

and

$$\hat{X}_t(m+1) = \psi_{m+1}\epsilon_t + \psi_{m+2}\epsilon_{t-1} + \psi_{m+3}\epsilon_{t-2} + \cdots.$$

Subtracting, we have

$$\hat{X}_{t+1}(m) - \hat{X}_t(m+1) = \psi_m \epsilon_{t+1}$$

or

$$\hat{X}_{t+1}(m) = \hat{X}_t(m+1) + \psi_m \epsilon_{t+1}.$$

Notice that $\hat{X}_{t+1}(m)$ is an updated forecast of $X_{t+m+1}$ with the shift in origin from t to t + 1. $\hat{X}_{t+1}(m)$ and $\hat{X}_t(m+1)$ both forecast $X_{t+m+1}$. Thus to update a forecast, we only need to add a single term, $\psi_m \epsilon_{t+1}$, to the old forecast. Recall also that $\epsilon_{t+1} = X_{t+1} - \hat{X}_t(1)$. Thus the update term is

$$\psi_m[X_{t+1} - \hat{X}_t(1)].$$

To calculate the $\psi$-weights, we use the same standard device as in the previous section

$$\Phi(B)\psi(B) = \theta(B)$$

so that

$$(I - \Phi_1 B - \cdots - \Phi_{p+d})(I + \psi_1 B + \psi_2 B^2 + \cdots) = (I - \theta_1 B - \cdots - \theta_q B^q).$$

Equating coefficients, as before,

$$\begin{aligned}
\psi_1 &= \Phi_1 - \theta_1 \\
\psi_2 &= \Phi_1 \psi_1 + \Phi_2 - \theta_2 \\
&\vdots \\
\psi_j &= \Phi_1 \psi_{j-1} + \cdots + \Phi_{p+d} \psi_{j-p-d} - \theta_j
\end{aligned}$$

where $\psi_0 = 1$, $\psi_j = 0$, $j < 0$ and $\theta_j = 0$, $j > q$.

For example, in our example in the last chapter, we were considering $(I - .8B)(I - B)X_t = \epsilon_t$. The computation thus is

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= 1.8 \\
\psi_2 &= 1.8\ \psi_1 - .8 \\
&\vdots \\
\psi_j &= 1.8\ \psi_{j-1} - .8\ \psi_{j-2}.
\end{aligned}$$

From this it is simple to compute, $\psi_0 = 1$, $\psi_1 = 1.8$, $\psi_2 = 2.44$, $\psi_3 = 2.95$, $\psi_4 = 3.38$, ... . From this it is a simple step to update our forecasts. Notice, that the $\psi$-weights are increasing. This is as it should be since this is a nonstationary model and the $\psi$-weights will not converge.

**13.5 Confidence Limits for Forecasts** Finally, we would like to address the question of confidence limits for the forecasts. Let us consider $X_{t+m}$. We first observe that $E[X_{t+m} \mid t] = \hat{X}_t(m)$ so that

$$\text{var}[X_{t+m} \mid t] = E[\{X_{t+m} - \hat{X}_t(m)\}^2 \mid t] = E[\{e_t(m)\}^2 \mid t] = (1 + \psi_1^2 + \cdots + \psi_{m-1}^2)\sigma_\epsilon^2.$$

Assuming the $\epsilon_t$ are normally distributed, then so is $X_{t+m} - \hat{X}_t(m)$. Thus upper and lower probability limits for $X_{t+m}$ are

$$\hat{X}_t(m) + p_{\frac{\alpha}{2}} \left(1 + \sum_{j=1}^{m+1} \psi_j^2\right)^{\frac{1}{2}} \sigma_\epsilon$$

and

$$\hat{X}_t(m) - p_{\frac{\alpha}{2}}\left(1 + \sum_{j=1}^{m+1} \psi_j^2\right)^{\frac{1}{2}} \sigma_\epsilon$$

where $p_{\frac{\alpha}{2}}$ is the $(1 - \alpha/2) \times 100$ percentile of a standard normal distribution. If we know the model exactly, then, of course, $\sigma_\epsilon$, is known.

In closing this section, it is worth noting that all of this chapter presumes that we know the model exactly. Thus, none of the coefficients $\Phi_j$, $\theta_j$, $\pi_j$ or $\psi_j$ are estimated. None of this theory really works if when we take the conditional expectations, $E[\cdot \mid t]$, we must account for coefficients $\Phi_j$, $\theta_j$, $\pi_j$ and $\psi_j$ which are random variables depending on the observations, $X_1, \ldots, X_t$. In practice, of course, we use estimates of $\Phi_j$, $\theta_j$, $\pi_j$ and $\psi_j$ based on the data and then forecast into the future based on these estimated coefficients. This is strictly speaking a questionable practice from a theoretical point of view and is truly a weakness of the theory just developed. The situation is improved somewhat is the data used to estimate the coefficients, $\Phi_j$, $\theta_j$, $\pi_j$ and $\psi_j$ is different from the data used for forecasting $X_{t+m}$. In this case the conditional expectations are at least legitimate and the minimum mean square error forecast holds at least conditionally. Unfortunately, this is often not possible, for example, with economic forecasting. In any case, the procedures described above seem to have value even though the proofs given may not be entirely legitimate in the setting in which these procedures are often used.

# Appendix A

# Riemann-Stieltjes Integrals

In our discussion of random process models in Chapter 6, it is convenient to use the concept of Riemann-Stieltjes integration. Consider an interval $(\alpha, \beta)$ and let $x_o < x_1 < \cdots < x_n$ be a partition of $(\alpha, \beta)$ so that

(A.1) $$\alpha = x_0 < x_1 < \cdots < x_n = \beta$$

Let $\Delta x = \max_j (x_j - x_{j-1})$ and consider the sum

(A.2) $$\sum_{j=1}^{n} g(x_j)\,(F(x_j) - F(x_{j-1})).$$

The Riemann-Stieltjes integral is defined as

(A.3) $$\int_{\alpha}^{\beta} g(x)\, d\,F(x) = \lim_{\substack{n \to \infty \\ \Delta x \to 0}} \sum_{j=1}^{n} g(x_j)(F(x_j) - F(x_{j-1}))$$

If F is the identity function, the Riemann-Stieltjes integral is the ordinary integral as defined in elementary calculus. The function F in general must be of bounded variation, but for our purposes, we will make the more restrictive assumption that F is non-decreasing. The usual properties of integrals hold.

Property 1. $\int_{\alpha}^{\beta} (g(x) + h(x))\, dF(x) = \int_{\alpha}^{\beta} g(x)\, dF(x) + \int_{\alpha}^{\beta} h(x)\, dF(x)$.

Property 2. $\int_{\alpha}^{\beta} dF(x) = F(\beta) - F(\alpha)$.

Property 3. If $f(x) = F'(x) = \frac{d\,F(x)}{dx}$,

then $\int_{\alpha}^{\beta} g(x)\, dF(x) = \int_{\alpha}^{\beta} g(x)\, f(x)\, dx$.

Property 4. If $g(x) = 1$, $\int_{\alpha}^{\beta} f(x)\, dx = F(\beta) - F(\alpha)$

and, hence,

$$F(x) = \int_{-\infty}^{x} f(u)\, du.$$

If F is a function of two variables, x, y, and

$$\alpha = x_0 < x_1 < \cdots < x_n = \beta \text{ and } \gamma = y_0 < y_1 < \cdots < y_m = \delta$$

and also $\Delta x = \max (x_j - x_{j-1})$, $\Delta y = \max (y_j - y_{j-1})$, then

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0 \\ n,m \to \infty}} \sum_{i=1}^{n} \sum_{j=1}^{m} g(x_i, y_j) \left(F(x_i, y_j) - F(x_i, y_{j-1}) - F(x_{i-1}, y_j) + F(x_{i-1}, y_{j-1})\right)$$

$$= \int_{\alpha}^{\beta} \int_{\gamma}^{\delta} g(x, y) \; F(dx, dy).$$

WOLFER SUNSPOT NUMBERS: YEARLY*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1770 | 101 | 1795 | 21 | 1820 | 16 | 1845 | 40 |
| 1771 | 82 | 1796 | 16 | 1821 | 7 | 1846 | 62 |
| 1772 | 66 | 1797 | 6 | 1822 | 4 | 1847 | 98 |
| 1773 | 35 | 1798 | 4 | 1823 | 2 | 1948 | 124 |
| 1774 | 31 | 1799 | 7 | 1824 | 8 | 1949 | 96 |
| 1775 | 7 | 1800 | 14 | 1825 | 17 | 1850 | 66 |
| 1776 | 20 | 1801 | 34 | 1826 | 36 | 1851 | 64 |
| 1777 | 92 | 1802 | 45 | 1827 | 50 | 1852 | 54 |
| 1778 | 154 | 1803 | 43 | 1828 | 62 | 1953 | 39 |
| 1779 | 125 | 1804 | 48 | 1829 | 67 | 1954 | 21 |
| 1780 | 85 | 1805 | 42 | 1830 | 71 | 1855 | 7 |
| 1781 | 68 | 1806 | 28 | 1831 | 48 | 1856 | 4 |
| 1782 | 38 | 1807 | 10 | 1832 | 28 | 1857 | 23 |
| 1783 | 23 | 1808 | 8 | 1833 | 8 | 1858 | 55 |
| 1784 | 10 | 1809 | 2 | 1834 | 13 | 1859 | 94 |
| 1785 | 24 | 1810 | 0 | 1835 | 57 | 1860 | 96 |
| 1786 | 83 | 1811 | 1 | 1836 | 122 | 1861 | 77 |
| 1787 | 132 | 1812 | 5 | 1837 | 138 | 1862 | 59 |
| 1788 | 131 | 1813 | 12 | 1838 | 103 | 1863 | 44 |
| 1789 | 118 | 1814 | 14 | 1839 | 86 | 1864 | 47 |
| 1790 | 90 | 1815 | 35 | 1840 | 63 | 1865 | 30 |
| 1791 | 67 | 1816 | 46 | 1841 | 37 | 1866 | 16 |
| 1792 | 60 | 1817 | 41 | 1842 | 24 | 1867 | 7 |
| 1793 | 47 | 1818 | 30 | 1843 | 11 | 1868 | 37 |
| 1794 | 41 | 1819 | 24 | 1844 | 15 | 1869 | 74 |

*100 Observations