NOTES ON STOCHASTIC DIFFERENCE EQUATIONS

Thomas Sargent

February 1977

Working Paper #: 66

Rsch. File #: 250.1

The views expressed herein are solely those of the author and do not necessarily represent the views of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

Contents

Stochastic Difference Equations

Introduction

Deterministic (nonrandom) difference equations of low order can generate "cycles," but not of the kind ordinarily thought to characterize economic variables. For example, we have seen that second-order difference equations can generate cycles of constant periodicity that are damped, explosive, or, in the very special case where the amplitude $r=1$, of constant-amplitude. But the "cycles" in economic variables seem neither damped nor explosive, and they don't have a constant period from one cycle to the next; e.g., some recessions last one year, some last for one and a half years. The "business cycle" is the tendency of certain economic variables to possess persistent cycles of approximately constant amplitude and somewhat irregular periodicity from one "cycle" to the other. The distinguishing characteristic of "the" business cycle is the apparent tendency of a number of important aggregate economic variables to move together, with timing relationships among the variables that tend to remain the same from one expansion-recession cycle to another. The National Bureau of Economic Research has inspected masses of data that indicate the presence of a business cycle of average length of about three years from peak to peak in many important economic aggregates for the U.S. The Bureau has also documented the tendency for the timing relationships among variables to remain somewhat the same from cycle to cycle.

Figure 1 graphs the 91-day Treasury Bill rate and the unemployment rate over the postwar period for quarterly data. The "business cycle" shows up in both series, interest rates tending to be high and unemployment low in "booms," and interest rates tending to be low and unemployment high in recessions. Clearly the "cycles" are irregular in length and don't "look like" those generated by our low-order difference equations.

As we have seen, low-order nonstochastic difference equations do not generate data that look as irregular as do the graphs of economic data just illustrated. However, high-order nonstochastic difference equations can generate data that look like economic data. For example, if y_t is governed by a nonstochastic n^{th} order homogeneous difference equation, its solution can be written

$$(1) \qquad y_t = \sum_{j=1}^{n} a_j \lambda_j^t$$

where the λ_j's are the roots of the characteristic equation and the a_j's are chosen to satisfy n initial conditions. By making n large enough, any sample of data can be modeled arbitrarily well with the nonstochastic equation (1). However, this device of using high-order nonrandom difference equations is generally regarded as an unpromising one for two reasons. First, to get a model that is capable of generating time series that resemble economic data well, the order of the difference equation must be made quite large, so that the model is not parsimonious in terms of its parameterization. Second, strictly speaking, the model (1) implies that once the appropriate equation is fit, perfect predictions of the future of y can be made. Most economists believe that predictions will always be subject to error, so that it seems advisable to adopt a model that recognizes this condition.

While low-order nonrandom difference equations don't provide an adequate model for explaining the cycles in economic data, low-order stochastic or random difference equations do. In effect, if the initial conditions of low-order deterministic difference equations are subjected to repeated random shocks of a certain kind, there emerges the possibility of recurring, somewhat irregular cycles of the kind seemingly infesting economic data. This is an important idea that is really the foundation of macroeconometric models, an idea that was introduced into economics by Slutsky and Frisch. These pages describe the elements of stochastic difference equations and some of their applications in macroeconomics.*

Preliminary Concepts

A stochastic process is a collection of random variables, a collection indexed by a variable t. In our work, we will regard t as time and will require t to be an integer, so that we'll be working in discrete time. Thus, the stochastic process y_t is a collection of random variables ... y_{-1}, y_0, y_1, y_2, ..., there being one random variable for each point in time t belonging to the set T, which in our case is the set of integers. Alternatively, on each "drawing," we draw an entire sequence $\{y_k\}_{k=-\infty}^{\infty}$. We are interested in the probability distribution of such sequences. A single drawing of a sequence $\{y_k\}$ is called a realization of the stochastic process y_t.

We will characterize the probability law governing the collection of random variables that make up the stochastic process by the list of means of y_t and by the covariances between y's at different points in time. (For a stochastic process that obeys the normal probability law, these parameters completely characterize the probability distribution. Even where y isn't normal, the first and second moments contain much

---

* The reader is assumed to be familiar with complex variables. The chapter on complex variables in R.G.D. Allen's Mathematical Economics is a good reference.

useful information, enough information to characterize the linear structure of the process.) In particular, we have that the mean of the process y_t is

$$Ey_t = \mu_t \qquad t \varepsilon T .$$

where E is the mathematical expectation operator. The covariances are given by

$$E[(y_t-\mu_t)(y_s-\mu_s)] = \sigma_{t,s} .$$

A stochastic process is said to be wide-sense <u>stationary</u> (or covariance stationary or second-order stationary) if μ_t is independent of t and if $\sigma_{t,s}$ depends only on t-s. We will henceforth deal with such stationary processes. The first and second moments of a stationary process are summarized by the mean μ and the <u>covariogram</u> $c(\tau)$ defined by

$$E[(y_t-\mu)(y_s-\mu)] = \sigma_{t,s}$$

$$= E[(y_t-\mu)(y_{t-\tau}-\mu)] = \sigma_{t,t-\tau} \equiv c(\tau) ,$$

where τ=t-s. The covariogram is easily verified to be symmetric, i.e., $c(\tau)=c(-\tau)$, and to obey $c(0)\geq|c(\tau)|$ for all τ, this inequality being an implication of the Schwarz inequality.

To find further restrictions on the covariogram, let x_t be a covariance stationary stochastic process with mean zero and covariogram $c(\tau)$. Consider forming a weighted sum of x's at different dates

$$y = \sum_{j=1}^{n} a_j x_{t_j}$$

where the a_j's are fixed real numbers and $t_1, \ldots, t_n$ are integers. We must require that the random variable y have nonnegative variance, so that

$$Ey^2 = E\left(\sum_{j=1}^{n} a_j x_{t_j} \sum_{k=1}^{n} a_k x_{t_k}\right)$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k E x_{t_k} x_{t_j}$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k c(t_k - t_j) \geq 0 \quad .$$

This last inequality is required to hold for any n, any list of a_j's, and any selection of $(t_1, t_2, \ldots, t_n)$. A sequence $c(\tau)$ that satisfies this condition is said to be "nonnegative definite." The condition that $c(\tau)$ be nonnegative definite is a necessary and sufficient condition for a sequence $c(\tau)$ to be the covariogram of a well-defined stochastic process.*

A basic building block is the serially uncorrelated random process ε_t, which satisfies

(2)
$$E(\varepsilon_t) = 0 \qquad \text{for all } t$$
$$E(\varepsilon_t^2) = \sigma_\varepsilon^2 \qquad \text{for all } t$$
$$E(\varepsilon_t \varepsilon_{t-s}) = 0 \qquad \text{all } t \text{ and all } s \neq 0 \quad .$$

This process is (wide-sense) stationary, each variate being uncorrelated with itself lagged $s = \pm 1, \pm 2, \ldots$ times, and is said to be serially uncorrelated. The process is also often referred to as "white noise." As we shall see, such a white-noise process can be viewed as the basic building block for a large class of stationary stochastic processes.

---

* The condition turns out to be equivalent with the condition that the spectral density of x be nonnegative, a condition which also in effect stems from the requirement that the variance of every linear combination of x's at different points in time be nonnegative.

To illustrate how the white-noise process ε_t can be used to build up more complicated processes, consider the random process y_t

$$y_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}$$

$$= B(L)\varepsilon_t$$

where $B(L) = \sum_{j=0}^{\infty} b_j L^j$, and where we assume $\sum_{j=0}^{\infty} b_j^2 < \infty$, a requirement needed to assure that the variance of y is finite. We assume that the ε process is "white" and thus satisfies properties (2). Equation (2) says that the y process is a one-sided moving sum of a white noise process, ε.

We seek the covariogram of the y process, i.e., we seek the values of $c_y(k)=E(y_t y_{t-k})$ for all k. It will be convenient to obtain the <u>covariance generating function</u> $g_y(z)$ which is defined by

(3) $$g_y(z) = \sum_{k=-\infty}^{\infty} c_y(k) z^k.$$

The coefficient on z^k in (3) is the k^{th} lagged covariance, $c_y(k)$.

First notice that taking mathematical expectations on both sides of (2) gives

$$E(y_t) = \sum_{j=0}^{\infty} b_j E(\varepsilon_{t-j})$$

$$= 0 \qquad \text{for all } t.$$

It therefore follows that

$$c_y(k) = E\{(y_t - Ey_t)(y_{t-k} - Ey_{t-k})\}$$

$$= Ey_t \cdot y_{t-k} \qquad \text{for all } k.$$

Notice $y_t \cdot y_{t-k}$ is

$$y_t y_{t-k} = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} \sum_{h=0}^{\infty} b_h \varepsilon_{t-k-h}$$

$$= (b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \ldots)(b_0 \varepsilon_{t-k} + b_1 \varepsilon_{t-k-1} + b_2 \varepsilon_{t-k-2} + \ldots)$$

$$y_t y_{t-k} = \{b_0 b_k \varepsilon_{t-k}^2 + b_1 b_{k+1} \varepsilon_{t-k-1}^2 + b_2 b_{k+2} \varepsilon_{t-k-2}^2 + \ldots\}$$

+ crossproduct terms whose expectations are zero.

Thus

$$c_y(k) = Ey_t y_{t-k} = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} b_j b_{j+k}. \tag{4}$$

The covariance generating function is then

$$g_y(z) = \sum_{k=-\infty}^{\infty} z^k c_y(k)$$

$$= \sigma_\varepsilon^2 \sum_{k=-\infty}^{\infty} z^k \sum_{j=0}^{\infty} b_j b_{j+k}$$

$$= \sigma_\varepsilon^2 \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} b_j b_{j+k} z^k$$

$$g_y(z) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} b_j b_{j+k} z^k.$$

Let $h=j+k$, so that $k=h-j$. Writing the above line in terms of the index h then gives

$$g_y(z) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \sum_{h=0}^{\infty} b_j b_h z^{h-j}$$

$$= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} b_j z^{-j} \sum_{h=0}^{\infty} b_h z^h.$$

The last equation gives the convenient expression

(5) $$g_y(z) = \sigma_\varepsilon^2 B(z^{-1})B(z)$$

where $B(z^{-1}) = \sum_{j=0}^{\infty} b_j z^{-j}$, $B(z) = \sum_{j=0}^{\infty} b_j z^j$.

Equation (5) gives the covariance generating function $g_y(z)$ in terms of the b_j's and the variance σ_ε^2 of the white noise ε.

To take an example that illustrates the usefulness of (5), consider the first-order process

(6) $$y_t = \lambda y_{t-1} + \varepsilon_t$$

or

$$y_t = \left(\frac{1}{1-\lambda L}\right)\varepsilon_t = \sum_{i=0}^{\infty} \lambda^i \varepsilon_{t-i}, \qquad |\lambda| < 1$$

where, as always, ε is a white-noise process with variance σ_ε^2. We have

$$B(L) = \frac{1}{1-\lambda L},$$

$$B(z) = \frac{1}{1-\lambda z} = 1 + \lambda z + \lambda^2 z^2 + \dots$$

$$B(z^{-1}) = \frac{1}{1-\lambda z^{-1}} = 1 + \lambda z^{-1} + \lambda^2 z^{-2} + \dots.$$

(Thus, $B(z)$ is found by replacing L in $B(L)$ by z.) So applying (5), we have

(7) $$g_y(z) = \sigma_\varepsilon^2 \left(\frac{1}{1-\lambda z^{-1}}\right)\left(\frac{1}{1-\lambda z}\right).$$

From our experience with difference equations we know that the expression (7) can be written as a sum

$$(8) \qquad g_y(z) = \frac{k_1\sigma_\varepsilon^2}{1-\lambda z} + \frac{k_2\sigma_\varepsilon^2 z^{-1}}{1-\lambda z^{-1}}$$

where k_1 and k_2 are certain constants. To find out what the constants must be, notice that (8) implies

$$g_y(z) = \sigma_\varepsilon^2 k_1(1+\lambda z+\lambda^2 z^2+\ldots)$$

$$+ \sigma_\varepsilon^2 k_2(z^{-1}+\lambda z^{-2}+\lambda^2 z^{-3}+\ldots),$$

so that $c_y(0)=k_1\sigma_\varepsilon^2$ and $c_y(1)=\sigma_\varepsilon^2\lambda k_1=\sigma_\varepsilon^2 k_2$.

By direct computation using (6) we note that

$$Ey_t^2 = \sum_{i=0}^{\infty} \lambda^{i2} E\varepsilon_t^2 = \frac{\sigma_\varepsilon^2}{1-\lambda^2}$$

$$Ey_t y_{t-1} = E\sum_{i=0}^{\infty} \lambda^i \varepsilon_{t-i} \sum_{i=1} \lambda^{i-1}\varepsilon_{t-i} = E\sum_{i=1} \lambda^i \lambda^{i-1} \varepsilon_{t-i}^2$$

$$= \sigma_\varepsilon^2 \lambda \sum_{i=1}^{\infty} \lambda^{2(i-1)} = \frac{\lambda\sigma_\varepsilon^2}{1-\lambda^2} \; .$$

So for (8) to be correct, we require that

$$k_1 = \frac{1}{1-\lambda^2}$$

$$k_2 = \frac{\lambda}{1-\lambda^2} \; .$$

With these values of k_1 and k_2, we can verify directly that

$$\sigma_\varepsilon^2 \quad \frac{\frac{1}{1-\lambda^2}}{1-\lambda z} + \frac{z^{-1}(\frac{\lambda}{1-\lambda^2})}{1-\lambda z^{-1}}$$

$$= \sigma_\varepsilon^2 \cdot \frac{1}{1-\lambda^2} \left[\frac{(1-\lambda z^{-1})+\lambda z^{-1}-\lambda^2}{(1-\lambda z)(1-\lambda z^{-1})}\right]$$

$$= \sigma_\varepsilon^2 \frac{1}{(1-\lambda z)(1-\lambda z^{-1})} ,$$

so that (8) and (7) are equivalent.

Expression (8) is the more convenient of the two expressions since it yields quite directly

$$g_y(z) = \sigma_\varepsilon^2 \frac{1}{1-\lambda^2} \left[\frac{1}{1-\lambda z} - \frac{\lambda z^{-1}}{1-\lambda z^{-1}}\right]$$

(9) $$= \sigma_\varepsilon^2 \frac{1}{1-\lambda^2} [\{1+\lambda z+\lambda^2 z^2+\dots\} + \{\lambda z^{-1}+\lambda^2 z^{-2}+\lambda^3 z^{-3}+\dots\}].$$

Thus, we have that for the "geometric" process (6),

$$c_y(k) = \frac{\sigma_\varepsilon^2}{1-\lambda^2} \cdot \lambda^{|k|} \qquad k=0, \pm 1, \pm 2, \dots$$

The covariance declines geometrically with increases in $|k|$. We require $|\lambda|<1$ in order that the y process have a finite variance.

To get this result more directly write the stochastic difference equation $y_t=\lambda y_{t-k}+\varepsilon_t$, then multiply y_t by y_{t-k}, $k>0$, to obtain

$$y_t y_{t-k} = \lambda y_{t-1} y_{t-k} + \varepsilon_t y_{t-k}.$$

Taking expected values on both sides and noting that $E\varepsilon_t y_{t-k}=0$ gives the famous Yule-Walker equation,

$$E(y_t y_{t-k}) = \lambda E(y_{t-1} y_{t-k})$$

or

$$c_y(k) = \lambda c_y(k-1) \qquad k > 0$$

which implies the solution

$$c_y(k) = \lambda^k c_y(0) \qquad\qquad k > 0$$

From the symmetry of covariograms, it then follows that $c_y(k) = \lambda^{|k|} c_y(0)$ for all k. Notice that the covariogram obeys the solution of the non-random part of the difference equation with initial condition $c_y(0)$.

As a second example, consider the second-order process

(10) $$y_t = \left(\frac{1}{1-\lambda_1 L}\right)\left(\frac{1}{1-\lambda_2 L}\right)\varepsilon_t, \qquad |\lambda_1+\lambda_2| < 1, \ \lambda_1 \neq \lambda_2,$$

where ε_t is white noise with variance σ_ε^2.

For (10) we have

$$B(L) = \left(\frac{1}{1-\lambda_1 L}\right)\left(\frac{1}{1-\lambda_2 L}\right)$$

$$B(z) = \left(\frac{1}{1-\lambda_1 z}\right)\left(\frac{1}{1-\lambda_2 z}\right)$$

$$B(z^{-1}) = \left(\frac{1}{1-\lambda_1 z^{-1}}\right)\left(\frac{1}{1-\lambda_2 z^{-1}}\right).$$

Applying formula (5), we have that the covariance-generating function is

(11) $$g_y(z) = \sigma_\varepsilon^2 \frac{1}{(1-\lambda_1 z)} \frac{1}{(1-\lambda_2 z)} \frac{1}{(1-\lambda_1 z^{-1})} \frac{1}{(1-\lambda_2 z^{-1})}.$$

Notice that (10) can be written

$$y_t = \frac{\lambda_1}{\lambda_1-\lambda_2}\left(\frac{1}{1-\lambda_1 L}\right)\varepsilon_t - \left(\frac{\lambda_2}{\lambda_1-\lambda_2}\right)\left(\frac{1}{1-\lambda_2 L}\right)\varepsilon_t$$

(12) $$y_t = \frac{\lambda_1}{\lambda_1-\lambda_2}\sum_{i=0}^{\infty}\lambda_1^i \varepsilon_{t-i} - \frac{\lambda_2}{\lambda_1-\lambda_2}\sum_{i=0}^{\infty}\lambda_2^i \varepsilon_{t-i}.$$

For y_{t-k}, $k \geq 0$, we have

(13) $$y_{t-k} = \frac{\lambda_1}{\lambda_1-\lambda_2} \sum_{i=k}^{\infty} \lambda_1^{i-k} \varepsilon_{t-i} - \frac{\lambda_2}{\lambda_1-\lambda_2} \sum_{i=k}^{\infty} \lambda_2^{i-k} \varepsilon_{t-i}.$$

Multiplying (12) and (13) together and taking expectations gives

$$E(y_t y_{t-k}) = \sigma_\varepsilon^2 \{\frac{\lambda_1^2}{(\lambda_1-\lambda_2)^2} \sum_{i=0}^{\infty} \lambda_1^{k+i} \lambda_1^i + \frac{\lambda_2^2}{(\lambda_1-\lambda_2)^2} \sum_{i=0}^{\infty} \lambda_2^{k+i} \lambda_2^i$$

$$- \frac{\lambda_1\lambda_2}{(\lambda_1-\lambda_2)^2} \sum_{i=0}^{\infty} \lambda_1^{k+i} \lambda_2^i - \frac{\lambda_1\lambda_2}{(\lambda_1-\lambda_2)^2} \sum_{i=0}^{\infty} \lambda_2^{k+i} \lambda_1^i\}$$

(14) $$Ey_t y_{t-k} = (\frac{1}{\lambda_1-\lambda_2})^2 \sigma_\varepsilon^2 [\frac{\lambda_1^{2+k}}{(1-\lambda_1^2)} + \frac{\lambda_2^{2+k}}{(1-\lambda_2^2)} - \frac{\lambda_1\lambda_2}{1-\lambda_1\lambda_2} (\lambda_1^k+\lambda_2^k)]$$

$k \geq 0.$

So (14) and the symmetry of $g_y(z)$ suggests that the appropriate factorization of (11) is

(15) $$g_y(z) = (\frac{1}{\lambda_1-\lambda_2})^2 \sigma_\varepsilon^2\{(\frac{\lambda_1^2}{(1-\lambda_1^2)} - \frac{\lambda_1\lambda_2}{1-\lambda_1\lambda_2})(\frac{1}{1-\lambda_1 z} + \frac{\lambda_1 z^{-1}}{1-\lambda_1 z^{-1}})$$

$$+ (\frac{\lambda_2^2}{1-\lambda_2^2}) - \frac{\lambda_1\lambda_2}{1-\lambda_1\lambda_2})(\frac{1}{1-\lambda_2 z} + \frac{\lambda_2 z^{-1}}{1-\lambda_2 z^{-1}})\} .$$

According to (14) and (15) the covariogram of a y process governed by the second-order process (10) consists of a weighted sum of two geometric decay processes, the decay parameters being λ_1 and λ_2, the inverse zeroes of the polynomial $(1-\lambda_1 L)(1-\lambda_2 L)$. Expression (14) implies that the covariogram displays damped oscillations if the roots λ_1 and λ_2 are complex conjugates. This can be shown by substituting $\lambda_1=re^{-iw}$ and $\lambda_2=re^{iw}$ into (14), and proceeding to analyze (14) as we above analyzed the solution of the deterministic (nonrandom) second-order difference equation. An alternative way to reach the same conclusion is as follows. Multiply both sides of (10) by $(1-\lambda_1 L)(1-\lambda_2 L)$ to get

(16) $$y_t = t_1 y_{t-1} + t_2 y_{t-2} + \varepsilon_t$$

where $t_1=(\lambda_1+\lambda_2)$ and $t_2=-\lambda_1\lambda_2$. Multiply (16) by y_{t-k} for $k\geq 0$ to get

$$y_t y_{t-k} = t_1 y_{t-1} y_{t-k} + t_2 y_{t-2} y_{t-k} + \varepsilon_t y_{t-k}.$$

Since $E\varepsilon_t y_{t-k}=0$, we have

$$E(y_t y_{t-k}) = t_1 E(y_{t-1} y_{t-k}) + t_2 (y_{t-2} y_{t-k}) \qquad k \geq 0$$

which shows that $c_y(k)$ obeys the difference equation (the Yule-Walker equation)

(17) $$c_y(k) = t_1 c_y(k-1) + t_2 c_y(k-2).$$

So the covariogram of a second (n^{th}) order process obeys the solution to the deterministic second (n^{th}) order difference equation examined above. In particular, corresponding to (17) we consider the polynomial

(18) $$1 - t_1 k - t_2 k^2 = 0,$$

which has roots $1/\lambda_1$ and $1/\lambda_2$. (We know that $1-t_1k-t_2k$ equals $(1-\lambda_1 k)(1-\lambda_2 k)$, with roots $1/\lambda_1$ and $1/\lambda_2$.) Alternatively, multiply (18) by k^{-2} to obtain

$$k^{-2} - t_1 k^{-1} - t_2 = 0$$

(19) $$x^2 - t_1 x - t_2 = 0 \text{ where } x = k^{-1}.$$

Notice that the roots of (19) are the reciprocals of the roots of (18), so λ_1 and λ_2 are the roots of (19).

The solution to the deterministic difference equation (17) is, as we have seen,

$$(20) \qquad c_y(k) = \lambda_1^k z_0 + \lambda_2^k z_1, \qquad k \geq 0$$

where z_0 and z_1 are certain constants chosen to make $c_y(0)$ and $c_y(1)$ equal the proper quantities. If the roots λ_1 and λ_2 are complex, we know from our work with deterministic difference equations that (20) becomes

$$(21) \qquad c_y(k) = z_0 \frac{r^k}{\sin w} \sin wk + z_1 \frac{r^k}{\sin w} \cos wk$$

where $\lambda_1 = re^{iw}$ and $\lambda_2 = re^{-iw}$. Accordingly to (21), the covariogram displays damped (we require $r<1$) oscillations with angular frequency w. A complete cycle occurs as wk goes from zero ($k=0$) to 2π ($k=2\pi/w$, if that is possible). The restrictions on t_1 and t_2 needed to deliver complex roots and so an oscillatory covariogram can be read directly from Figure _ of "Notes on Difference Equations."

Figure 4b below displays a realization of second-order processes for values of t_1 and t_2, values for which the roots are complex. Notice the tendency of these series to cycle, but with a periodicity that is somewhat variable from cycle to cycle.

The foregoing suggests one definition of a cycle in a single series: a series may be said to possess a "cycle" if its covariogram is characterized by (damped) oscillations. The typical "length" of the cycle can be measured by $2\pi/w$, where w is the angular frequency associated with the damped oscillations in the covariogram (e.g., see 21). To be labelled a business cycle the cycle should exceed a year in length. (Cycles of one year in length are termed "seasonals.")

The Cross Covariogram

Suppose we have two wide-sense stationary stochastic process y_t and x_t. The processes are said to be <u>jointly wide sense stationary</u> if the cross-covariance $E(y_t - Ey_t)(x_{t-k} - Ex_{t-k})$ depends only on k and not on t. The <u>cross-covariogram</u> is the list of these covariances viewed as a function of k. We denote it by

$$c_{yx}(k) = E(y_t - Ey)(x_{t-k} - Ex).$$

Now suppose that both y_t and x_t can be expressed as one-sided distributed lags of a single white-noise process ε_t:

$$y_t = B(L)\varepsilon_t$$

$$x_t = D(L)\varepsilon_t$$

where $B(L) = \sum_{j=0}^{\infty} b_j L^j$, $D(L) = \sum_{j=0}^{\infty} d_j L^j$. Since $E\varepsilon_t = 0$, we have

$$c_{yx}(k) = Ey_t x_{t-k}$$

$$= E \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} \sum_{h=0}^{\infty} d_h \varepsilon_{t-h-k}$$

$$= E(b_0\varepsilon_t + b\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \ldots)(d_0\varepsilon_{t-k} + d_1\varepsilon_{t-k-1} + d_2\varepsilon_{t-k-2} + \ldots)$$

$$c_{yx}(k) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} d_j b_{j+k}$$

The cross covariance generating function $g_{yx}(z)$ is defined by

$$g_{yx}(z) = \sum_{k=-\infty}^{\infty} c_{yx}(k) z^k,$$

the coefficient on z^k being $c_{yx}(k)$. In the present case, we have

$$g_{yx}(z) = \sigma_\varepsilon^2 \sum_{k=-\infty}^{\infty} z^k \sum_{j=0}^{\infty} d_j b_{j+k}$$

$$= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d_j b_{j+k} z^k$$

Letting h=j+k so that k=h-j, we have

$$g_{yx}(z) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \sum_{h=0}^{\infty} d_j b_k z^{h-j}$$

$$= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} d_j z^{-j} \sum_{j=0}^{\infty} b_h z^h$$

(22) $$g_{yx}(z) = \sigma_\varepsilon^2 D(z^{-1})B(z).$$

This is a counterpart to equation (5), and includes it as a special case.

Now suppose that we have the more general system

(23)
$$y_t = A(L)\varepsilon_t + B(L)u_t$$
$$x_t = C(L)\varepsilon_t + D(L)u_t$$

where ε_t and u_t are two mutually uncorrelated (at all lags) white noise processes with variances σ_ε^2 and σ_u^2 respectively, and $Eu_t\varepsilon_{t-k}=0$ for all k. By carrying out calculations analogous to those just completed, it is possible to express the cross-covariance generating function between y and x as

(24) $$g_{yx}(z) = \sigma_\varepsilon^2 A(z)C(z^{-1}) + \sigma_u^2 B(z)D(z^{-1})$$

As it turns out, (23) is a very general representation for a bivariate stochastic process, including a large class of such processes.*

A Mathematical Digression on
Fourier-Transforms and z-Transforms

The following theorem provides the foundation for the z-transform, Fourier-transform, and "lag operator" methods that we use repeatedly in these pages. The theorem, which we shall not prove,** is a version of the Riesz-Fisher theorem.

Theorem (Riesz-Fisher):

Let $\{c_n\}_{n=-\infty}^{\infty}$ be a sequence of complex numbers for which $\sum_{n=-\infty}^{\infty} |c_n|^2 < \infty$.

Then there exists a complex-valued function f(w) defined for real w's belonging to the interval $[-\pi,\pi]$, such that

(25) $$f(w) = \sum_{j=-\infty}^{\infty} c_j e^{-iwj} ,$$

where the infinite series converges in the "mean square" sense that

$$\lim_{n\to\infty} \int_{-\pi}^{\pi} \left| \sum_{j=-n}^{n} c_j e^{-iwj} - f(w) \right|^2 dw = 0 \quad .$$

The function f(w) is called the "Fourier transform of the c_k's" and satisfies

$$\left| \int_{-\pi}^{\pi} |f(w)|^2 dw \right| < \infty$$

---

* Namely, all jointly wide-sense stationary, indeterministic processes.

** For a proof of the Riesz-Fisher theorem, see Tom Apostol, Mathematical Analysis, second edition, Addison-Wesley, Chapter 11.

where the integral is a Lebesque integral (i.e. "f belongs to $L_2[-\pi,\pi]$"). Given f(w), the c_k's can be "recovered" from the inversion formula

(26) $$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(w)e^{+iwk}dw$$

Finally, the function f(w) and the c_k's satisfy Parseval's relation

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(w)|^2 dw = \sum_{j=-\infty}^{\infty}|c_j|^2$$

This completes the statement of the theorem.

Consider the space of all doubly infinite sequences $\{x_k\}_{k=-\infty}^{\infty}$ such that $\sum_{k=-\infty}^{\infty}|x_k|^2<\infty$, i.e., the space of square summable sequences. We denote this space $\ell_2(-\infty,\infty)$. It is a <u>linear space</u> in the sense that it possesses the following two properties (among others):

(i) Let α be a scalar and let $\{x_t\}$ belong to $\ell_2(-\infty,\infty)$. Then $\{\alpha x_k\}$ belongs to $\ell_2(-\infty,\infty)$, i.e. $\sum_{k=-\infty}^{\infty}|\alpha x_k|^2<\infty$.

(ii) Let $\{x_k\}$ and $\{y_k\}$ both belong to $\ell_2(-\infty,\infty)$. Then $\{x_k+y_k\}$ belongs to $\ell_2(-\infty,\infty)$, i.e. $\sum_{k=-\infty}^{\infty}(x_k+y_k)^2<\infty$.

Now consider the space $L_2[-\pi,\pi]$ consisting of all functions f(w) for which $\int_{-\pi}^{\pi}|f(w)|^2 dw<\infty$, i.e. the space of "square Lebesque integrable functions" on $[-\pi,\pi]$. We denote this space $L_2[\pi,\pi]$. This space is a <u>linear space</u> in the sense that it possesses the two properties:

(a) Let α be a scalar and let $f(w)$ belong to $L_2[-\pi,\pi]$. Then $\alpha f(w)$ belongs to $L_2[-\pi,\pi]$, i.e., $\int_{-\pi}^{\pi} |\alpha f(w)|^2 dw < \infty$.

(b) Let $f(w)$ and $g(w)$ both belong to $L_2[-\pi,\pi]$. Then $f(w)+g(w)$ belongs to $L_2[-\pi,\pi]$, i.e., $\int_{-\pi}^{\pi} |f(w)+g(w)|^2 dw < \infty$.

The spaces $\ell_2(-\infty,\infty)$ and $L_2[-\pi,\pi]$ are each metric spaces in the sense that each one possesses a well-defined metric or distance function. In particular, on $\ell_2(-\infty,\infty)$ the real valued function

$$d_2(x,y) = \left[\sum_{k=-\infty}^{\infty} (x_k - y_k)^2\right]^{\frac{1}{2}}$$

measures the distance between the two sequences $\{x_k\}$ and $\{y_k\}$. The function $d_2(\cdot,\cdot)$ is defined for all $\{x_k\}$ and $\{y_k\}$ in $\ell_2(-\infty,\infty)$ and is a "natural" measure of distance (it satisfies a triangle inequality $d(x,y) \leq d(x,z)+d(z,y)$ for all sequences x, y, and z in ℓ_2. On $L_2[-\pi,\pi]$ the real valued function

$$D_2(f,g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(w) - g(w)|^2 dw$$

is a metric that measures the "distance" between two functions $f(w)$ and $g(w)$. The metric $D_2(\cdot,\cdot)$ is defined for all $f(w)$ and $g(w)$ belonging to $L_2[-\pi,\pi]$.

Now consider the mapping from $\ell_2(-\infty,\infty)$ to $L_2[-\pi,\pi]$ defined by the Fourier transform

(25) $$f(w) = \sum_{k=-\infty}^{\infty} c_k e^{-iwk} \qquad w \epsilon [-\pi,\pi] \quad .$$

We also have the inverse mapping

(26) $$c_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(w) e^{+iwj} dw \qquad j=0,\pm 1,\pm 2,\ldots \quad .$$

Now a converse of the Riesz-Fisher theorem is also true: let $f(w)$ belong to $L_2[-\pi,\pi]$. Then there exists a sequence $\{c_k\}$ such that $\sum |c_k|^2 < \infty$ and

$$f(w) = \sum_{k=-\infty}^{\infty} c_k e^{-iwk}$$

where

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(w)e^{+iwk}dw \quad ,$$

where the infinite sum converges in the mean square sense. This converse theorem assures us that the mapping of $\ell_2(-\infty,\infty)$ into $L_2[-\pi,\pi]$ defined by (25) is onto. It is also one-to-one. The usefulness of the mapping (25) stems from the fact that it is an isometric isomorphism from $\ell_2(-\infty,\infty)$ to $L_2[-\pi,\pi]$; that is, it is a one-to-one and onto transformation of points in $\ell_2(-\infty,\infty)$ into points in $L_2[-\pi,\pi]$ that preserves both linear structures (i.e. it is an isomorphism) and distance between "points" (i.e. it is an isometric mapping). That is, let $\{x_k\},\{y_k\}$ belong to $\ell_2(-\infty,\infty)$, let α be a scalar, and let

$$x(w) = \sum_{k=-\infty}^{\infty} x_k e^{-iwk}$$

$$y(w) = \sum_{k=-\infty}^{\infty} y_k e^{-iwk}$$

Then we have (as can be verified directly)

$$x(w) + y(w) = \sum_{k=-\infty}^{\infty} (x_k + y_k)e^{-iwk}$$

$$x(w) = \sum_{k=-\infty}^{\infty} x_k e^{-iwk}$$

So "the Fourier transform of a sum of two sequences is the sum of their Fourier transforms" and "the Fourier transform of $\{\alpha x_k\}$ is α times the Fourier transform of $\{x_k\}$." This means that (25) is an isomorphism. We also have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|x(w) - y(w)|^2 dw = \left(\sum_{k=-\infty}^{\infty}(x_k - y_k)^2\right)^{\frac{1}{2}}$$

or

$$D_2(x(w),y(w)) = d_2(x,y)$$

so that (25) is an isometric mapping.

The Fourier transformation (25) puts square summable sequences $\{x_k\}$ into one-to-one correspondence with square integrable functions f(w) on $[-\pi,\pi]$. The transformation preserves linear structure and a measure of distance, as we have seen. The benefit from using the transformation is that operations that are complicated in one space are sometimes the counterparts of simple operations in another space. In particular, consider the <u>convolution</u> of two sequences $\{x_k\}$ and $\{y_k\}$ defined to be the new sequence

$$\{y * x_k\}_{k=-\infty}^{\infty} \equiv \sum_{j=-\infty}^{\infty} y_s x_{k-s}$$

The Fourier transform of $(y + x)_k$ is given by

$$\sum_{k=-\infty}^{\infty}\sum_{s=-\infty}^{\infty} y_s x_{k-s} e^{-iwk}$$

$$= \sum_{s=-\infty}^{\infty} y_s e^{-iws} \sum_{k=-\infty}^{\infty} x_{k-s} e^{-iw(k-s)}$$

$$= y(w)\cdot x(w)$$

where $y(w) = \sum_{k=-\infty}^{\infty} y_k e^{-iwk}$, $x(w) = \sum_{k=-\infty}^{\infty} x_k e^{-iwk}$.

Thus the Fourier transform of the <u>convolution</u> of $\{x_k\}$ with $\{y_k\}$ is the product of the Fourier transforms of $\{x_k\}$ and $\{y_k\}$. The complicated convolution operation corresponds simply to multiplication of Fourier transforms.

All transform techniques exploit properties like the preceding one. The aim is to transform a problem from one space where it appears complicated to another isometrically isomorphic space where the operations are simpler, then to transform back to the original space using the inversion mapping such as (26) after the calculations have been performed.

By making the change of variable $z=e^{-iw}$ in the Riesz-Fisher theorem, we obtain the following corollary which underlies our z-transform methods.

Corollary: Let $\{c_n\}_{n=-\infty}^{\infty}$ be a sequence of complex numbers for which $\sum_{n=-\infty}^{\infty} |c_n|^2 < \infty$. Then there exists a complex valued function g(z) with domain in the complex plain such that

$$g(z) = \sum_{j=-\infty}^{\infty} c_j z^j$$

where the infinite series converges in the mean square sense that

$$\lim_{n\to\infty} \int_\Gamma \left| \sum_{j=-n}^{n} c_j z^j - g(z) \right|^2 \frac{dz}{z} = 0$$

where Γ denotes the unit circle and the above integral is a contour integral. The function g(z) is defined at least on the unit circle in the complex plane and satisfies

$$\left| \frac{1}{2\pi i} \int_\Gamma |g(z)|^2 \frac{dz}{z} \right| < \infty \quad .$$

The function g(z) is called the "z-transform" of the sequence $\{c_k\}$. The c_k's can be recovered from g(z) by $c_k = \frac{1}{2\pi i} \int_\Gamma g(z) z^{-k-1} dz$. This completes the corollary.

So long as we restrict ourselves to sequences satisfying $\sum|c_k|^2 < \infty$, the theorem and the corollary guarantee that the "z-transforms" and Fourier transforms that we shall manipulate are well defined. The z-transform in effect maps the sequence $\{c_k\}$ into a complex-valued function defined on the unit circle in the complex plane. The Fourier transform maps the sequence $\{c_k\}$ into a complex-valued function defined on the real line over the interval $[-\pi,\pi]$.

Notice that the complex-valued functions e^{iwj}, $j=0, \pm 1, \pm 2, \ldots$ are an orthogonal set on the interval $[-\pi,\pi]$. That is, for $n \neq m$, we have

$$\frac{1}{2\pi}\int_{-\pi} e^{iwn}\cdot e^{-iwm}dw = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{iw(n-m)}dw$$

$$= \frac{1}{2\pi i(n-m)}\,[e^{iw(n-m)}]_{-\pi}^{\pi}$$

$$= \frac{1}{2\pi i(n-m)}\,[e^{i\pi(n-m)} - e^{-i\pi(n-m)}]$$

$$= \frac{1}{\pi(n-m)}\sin\pi(n-m) = 0$$

since $\sin\pi(n-m) = 0$ for $(n-m)$ an integer.

For the most part, the Riesz-Fisher theorem and its corollary are sufficient for our needs. Below we will briefly touch on a deterministic process for which the condition $\sum|c_k|^2 < \infty$ is violated (where the c_k's depict the covariogram) so that the theorem will not suffice to define the Fourier transform of the c_k's. It turns out that there is still a sense in which the Fourier transform of such "ill-behaved" $\{c_k\}$ sequences is defined, as we shall see.*

The Spectrum

An alternative representation of the covariance-generating function of y is the spectrum of the y process. Recall the covariance generating function of y defined in (3),

$$(3) \qquad g_y(z) = \sum_{k=-\infty}^{\infty} c_y(k) z^k .$$

For the process $y_t = B(L)\varepsilon_t$ we have seen that

$$g_y(z) = B(z)B(z^{-1})\sigma_\varepsilon^2 \quad .$$

If we evaluate (3) at the value $z = e^{-iw}$, we have

$$(27) \qquad g_y(e^{-iw}) = \sum_{k=-\infty}^{\infty} c_y(k) e^{-iwk} \qquad -\pi < w < \pi \quad .$$

Viewed as a function of angular frequency w, $g_y(e^{-iw})$ is called the <u>spectrum</u> of y. The spectrum is the Fourier transform of the covariogram.

As we would expect from the inversion formula (26), the spectrum is itself a kind of covariance generating function. Given an expression for $g_y(e^{-iw})$ it is easy to recover the covariances $g_y(k)$ from the inversion formula (26). To motivate the inversion formula, we multiply (27) by e^{iwh} and integrate with respect to w from $-\pi$ to π:

$$(28) \qquad \int_{-\pi}^{\pi} g_y(e^{-iw}) e^{iwh} dw = \int_{-\pi}^{\pi} \sum_{k=-\infty}^{\infty} c_y(k) e^{iw(h-k)} dw$$

$$= \sum_{k=-\infty}^{\infty} c_y(k) \int_{-\pi}^{\pi} e^{iw(h-k)} dw \quad .$$

Now for h=k we have

$$\int_{-\pi}^{\pi} e^{iw(h-k)}dw = \int_{-\pi}^{\pi} 1dw = 2\pi.$$

For h≠k we have,

$$\int_{-\pi}^{\pi} e^{iw(h-k)}dw = \int_{-\pi}^{\pi} \cos w(h-k)dw + i\int_{-\pi}^{\pi} \sin w(h-k)dw$$

$$= -\sin w(h-k)\Big]_{-\pi}^{\pi} + i \cos w(h-k)\Big]_{-\pi}^{\pi}$$

$$= 0.$$

Therefore, (28) becomes

$$\int_{-\pi}^{\pi} g_y(e^{-iw})e^{iwh}dw = 2\pi c_y(h).$$

Thus multiplying the spectrum by e^{iwh} and integrating from $-\pi$ to π gives the h^{th} lagged covariance times 2π. In particular, notice that for h=0, we have

$$\int_{-\pi}^{\pi} g_y(e^{-iw})dw = 2\pi c_y(0),$$

so that the area under the spectrum from $-\pi$ to π equals 2π times the variance of y. This fact motivates the interpretation of the spectrum as a device for decomposing the variance of a series by frequency. The portion of the variance of the series occurring between any two frequencies is given by the area under the spectrum between those two frequencies.

Notice that from (27) we have

$$g_y(e^{-iw}) = \sum_{k=-\infty}^{\infty} c_y(k)e^{-iwk}$$

(29) $$= c_y(0) + \sum_{k=1}^{\infty} c_y(k)(e^{iwk}+e^{-iwk})$$

$$= c_y(0) + 2\sum_{k=1}^{\infty} c_y(k) \cos wk.$$

According to (29) the spectrum is real valued at each frequency, and is obtained by multiplying the covariogram of y by a cosine function of the frequency in question. Notice also that since cos x=cos -x, it follows from (29) that

$$g_y(e^{iw}) = g_y(e^{-iw}),$$

so that the spectrum is symmetric about w=0.

Notice also that since cos $(w+2\pi k)$=cos (w), k=0, ± 1, ± 2, ..., it follows that the spectrum is a periodic function of w with period 2π. Therefore, we can confine our attention to the interval $[-\pi,\pi]$, or even $[0,\pi]$ by virtue of the symmetry of the spectrum about w=0.

We know that if

(30) $$y_t = B(L)\varepsilon_t,$$

where ε_t is white noise, then the spectrum of y is related to the spectrum of ε_t by

$$g_y(e^{-iw}) = B(e^{-iw})B(e^{iw})\sigma_\varepsilon^2$$

or

(31) $$g_y(e^{-iw}) = B(e^{-iw})B(e^{iw})g_\varepsilon(e^{-iw})$$

since for the white noise ε, $g_y(e^{-iw})=\sigma_\varepsilon^2$. It is straightforward to show that for any ε_t, not necessarily a white one, affecting y via (31), the spectrum of y is related to the spectrum of ε by (31). Thus, assume that y is related to X by

(32) $$y_t = \sum_{s=-p}^{q} b_s X_{t-s} \equiv B(L)X_t \qquad p \geq 0,\ q \geq 0$$

and that the spectrum of X is defined. From (32) we know that

$$y_t y_{t-j} = \sum_{s=-p}^{q} b_s X_{t-s} \sum_{r=-p}^{q} b_r X_{t-j-r}$$

$$= \sum_{s=-p}^{q} \sum_{r=-p}^{q} b_s b_r X_{t-s} X_{t-j-r}.$$

Taking expected values on both sides gives

$$c_y(j) = E(y_t y_{t-j}) = \sum_{s=-p}^{q} \sum_{r=-p}^{q} b_s b_r c_x(j+r-s).$$

The spectrum of y is defined as

$$g_y(e^{-iw}) = \sum_{k=-\infty}^{\infty} c_y(k) e^{-iwk}$$

(33) $$= \sum_{k=-\infty}^{\infty} \sum_{s=-p}^{\infty} \sum_{r=-p}^{\infty} b_r b_s c_x(k+r-s) e^{-iwk}.$$

Define the index h=k+r-s, so that k=h-r+s. Notice that

(34) $$e^{-iwk} = e^{-iw(h-r+s)} = e^{-iwh} e^{-iws} e^{iwr}.$$

Substituting (34) into (33) gives

$$g_y(e^{-iw}) = \sum_{r=-p}^{q} b_r e^{iwr} \sum_{s=-p}^{q} b_s e^{-iws} \sum_{h=-\infty}^{\infty} c_x(h) e^{-iwh}$$

(35) $$g_y(e^{iw}) = B(e^{iw}) B(e^{-iw}) g_x(e^{-iw})$$

or

(36) $$g_y(e^{-iw}) = |B(e^{iw})|^2 \; g_x(e^{-iw}),$$

which shows that the spectrum of the "output" y equals the spectrum of the "input" x multiplied by the positive real number $B(e^{iw})B(e^{-iw})$. Of course, it is also true that

$$g_y(z) = B(z)B(z^{-1})g_x(z).$$

Expression (36) motivates the interpretation of the spectrum as decomposing the variance of y by frequency. Thus, suppose we could choose $B(e^{-iw})$ so that

(37) $$B(e^{-iw}) = \begin{cases} 1 & \text{for} \quad w\varepsilon[a,b]\cap[-b,-a] \quad 0 < a < b < \pi. \\ 0 & \text{otherwise} \quad . \end{cases}$$

Thus, we are choosing a "filter," i.e., a set of b_j's, that takes a random process x_t and transforms it into a random process y_t according to (34). A filter obeying (37) shuts off all of the spectral power for frequencies not in the region [a,b] or [-b,-a]. To determine a set of b_j's that satisfies (37), we use the "inversion" formula seen earlier,

$$b_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} B(e^{-iw})e^{+iwj}dw$$

$$= \frac{1}{2\pi} \int_{-b}^{-a} e^{iwj}dw + \frac{1}{2\pi} \int_{a}^{b} e^{iwj}dw$$

$$= \frac{1}{2\pi} \int_{a}^{b} (e^{iwj}+e^{-iwj})dw$$

$$= \frac{1}{2\pi} \int_{a}^{b} 2 \cos wjdw$$

$$= \frac{1}{\pi}\frac{1}{j} \sin wj \Big]_{a}^{b}$$

(38) $$b_j = \frac{1}{\pi}\left(\frac{\sin jb - \sin ja}{j}\right), \text{ for all integer } j.$$

With the b_j's chosen in this way, the y process defined by

$$y_t = \sum_{j=-\infty}^{\infty} b_j x_{t-j}$$

has all of its variance occurring in the frequency bands w [a,b], w [-b,-a]. The variance of y is given by

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} g_y(e^{-iw})dw = \frac{1}{2\pi}\int_{-b}^{-a} g_x(e^{-iw})dw + \frac{1}{2\pi}\int_{a}^{b} g_x(e^{-iw})dw.$$

In this sense, $g_x(e^{-iw})$ gives a decomposition of the variance of x by frequency, the variance occurring over a given frequency being found by integrating the spectrum over that band and dividing by 2π. We have already seen that by integrating the spectrum from $-\pi$ to π we obtain the variance of x times 2π. As we shall show shortly, the decomposition of the variance of x by frequency that is reflected in the spectrum is one in which components at different frequencies can be regarded as orthogonal. More precisely two components formed by applying two filters like (37) that let through power over disjoint frequency bands are mutually orthogonal at all lags.

Incidentally, the preceding calculations can be used to prove that the spectrum is always nonnegative. This can be done by proceeding by contradiction. Suppose that the spectrum $g_x(e^{-iw})$ is negative over a small band. Then choose a filter that shuts off all variance outside of this band. The result is to produce a new random process that has a negative variance, a contradiction. So the spectrum must be nonnegative.

Let us examine the spectra of some simple processes. First consider the white noise process

$$y_t = \varepsilon_t$$

ε_t white so that $c_y(0)=\sigma_\varepsilon^2$, $c_y(h)=0$ for $h\neq 0$.

For this process the covariance generating function is simply

$$g_y(z) = \sigma_\varepsilon^2,$$

so that the spectrum is

$$g_y(e^{-iw}) = \sigma_\varepsilon^2, \qquad -\pi \leq w \leq \pi$$

so that the spectrum is flat, and equals σ_ε^2 at each frequency. Notice that

$$\int_{-\pi}^{\pi} g_y(e^{-iw})dw = 2\pi\sigma_\varepsilon^2,$$

as expected. So a white noise has a flat spectrum, indicating that all frequencies between $-\pi$ and π are equally important in accounting for its variance.

Next consider the first order process

$$y_t = B(L)\pi_t = \frac{1}{1-\lambda L}\varepsilon_t. \qquad -1 < \lambda < 1.$$

For this process the covariance generating function is

$$g_y(z) = (\frac{1}{1-\lambda z})(\frac{1}{1-\lambda z^{-1}})\sigma_\varepsilon^2.$$

Therefore, the spectrum is

$$g_y(e^{-iw}) = (\frac{1}{1-\lambda e^{-iw}})(\frac{1}{1-\lambda e^{iw}})\sigma_\varepsilon^2$$

$$= \frac{1}{(1-\lambda(e^{iw}+e^{-iw})+\lambda^2)}\cdot\sigma_\varepsilon^2$$

$$g_y(w) = \frac{1}{1-2\lambda\cos w+\lambda^2}\sigma_\varepsilon^2$$

Notice that

$$\frac{dg_y(w)}{dw} = -(1-2\lambda\cos w+\lambda^2)^{-2}(2\lambda\sin w)\sigma_\varepsilon^2$$

The first term in parenthesis is positive. Since $\sin w>0$ for $0<w<\pi$, the second term is negative on $(0,\pi)$ if $\lambda<0$ and positive on $(0,\pi]$ if $\lambda>0$. Therefore, if $\lambda>0$, the spectrum decreases on $(0,\pi]$ as w increases; if $\lambda<0$, the spectrum increases on $(0,\pi]$ as w increases. Thus, if $\lambda>0$, low frequencies (i.e., low values of w) are relatively important in composing the variance of w, while if $\lambda<0$, high frequencies are the more important. It is easy to verify that the higher in absolute value is λ, the steeper is the spectrum. Notice that the first order process can have a peak in its spectrum only at $w=0$ or $w=\pm\pi$. A peak at $w=\pi$ corresponds to a periodicity of $2\pi/w=2\pi/\pi=2$ periods. A peak at $w=0$, corresponds to a cycle with "infinite" periodicity, which is unobservable and hence not a cycle at all.

With quarterly data, a business cycle corresponds to a peak in the spectrum at a periodicity of about 12 quarters. A first-order process is capable of having a peak only at two quarters or at "infinite" quarters, and so is not capable of rationalizing a business cycle in the sense of a peak in the spectrum at about twelve quarters. As we saw above, a first-order process cannot possess a covariogram with a periodicity other than two periods, and so with quarterly data cannot rationalize a business cycle in the sense of an oscillatory covariogram.

Next consider the second-order process

$$y_t = \frac{1}{1-t_1L-t_2L^2} \varepsilon_t,$$

ε_t white noise. For this process the covariance generating function is

$$g_y(z) = \frac{1}{1-t_1z-t_2z^2} \frac{1}{1-t_1z^{-1}-t_2z^{-2}} \sigma_\varepsilon^2.$$

Therefore, the spectrum of the process is

$$g_y(e^{-iw}) = \frac{1}{1-t_1e^{-iw}-t_2e^{-2iw}} \frac{1}{1-t_1e^{iw}-t_2e^{2iw}} \sigma_\varepsilon^2$$

$$= \frac{\sigma_\varepsilon^2}{1+t_1^2+t_2^2+(t_2t_1-t_1)(e^{iw}+e^{-iw})-t_2(e^{-2iw}+e^{2iw})}$$

$$= \frac{\sigma_\varepsilon^2}{1+t_1^2+t_2^2-2t_1(1-t_2)\cos w-2t_2\cos 2w} = \frac{\sigma_\varepsilon^2}{h(w)} .$$

Differentiating with respect to w, we have

$$\frac{dg_y(e^{-iw})}{dw} = -\sigma_\varepsilon^2 h(w)^{-2}(2t_1(1-t_2)\sin w + 4t_2\sin 2w)$$

$$= -\sigma_\varepsilon^2 h(w)^{-2}(2\sin w\cdot[t_1(1-t_2)+4t_2\cos w]).$$

We know that $h(w)>0$. For the above derivative to be zero at a w belonging to $(0,\pi)$, we must have the term in brackets equal to zero:

$$t_1(1-t_2) + 4t_2 \cos w = 0$$

or

$$\cos w = \frac{-t_1(1-t_2)}{4t_2} \tag{39}$$

so that

(40) $$w = \cos^{-1}\left(\frac{-t_1(1-t_2)}{4t_2}\right) .$$

Equation (35) can be satisfied only if

(41) $$\left|\frac{-t_1(1-t_2)}{4t_2}\right| < 1,$$

since $|\cos x| \leq 1$ for all x. If (41) is met, the spectrum of y does achieve a maximum on $(0,\pi)$. Condition (41) is slightly more restrictve than the condition that the roots of the deterministic difference equation be complex so that the covariogram display oscillations. Let us write (41) as

(42) $$-1 < \frac{-t_1(1-t_2)}{4t_2} < 1.$$

The boundaries of the region (42) are

(43) $$-t_1(1-t_2) = 4t_2$$

and

(44) $$-t_1(1-t_2) = -4t_2.$$

The points $(t_1,t_2)=(0,0)$ appear on both boundaries, while the point $(t_1,t_2)=(2,-1)$ appears on (43) and $(t_1,t_2)=(-2,-1)$ appears on (44). Differentiating (43) implicitly with respect to t_1 gives

$$\frac{dt_2}{dt_1} = \frac{t_2-1}{4-t_1}$$

so that along (43)

$$\left.\frac{dt_2}{dt_1}\right)_{t_1=t_2=0} = -\frac{1}{4}$$

and

$$\left.\frac{dt_2}{dt_1}\right)_{\substack{t_1=2\\ t_2=-1}} = -1.$$

Differentiating (44) with respect to t_1 gives

$$\frac{dt_2}{dt_1} = \frac{1-t_2}{4+t_1}$$

so that along (44)

$$\left.\frac{dt_2}{dt_1}\right)_{t_1=t_2=0} = \frac{1}{4}$$

$$\left.\frac{dt_2}{dt_1}\right)_{\substack{t_1=-2\\ t_2=-1}} = 1.$$

Such calculations show that the boundaries of region (42) are as depicted in Figure 2. To be in region (42) with $t_2<1$ (a requirement of covariance stationarity) implies that the roots of the difference equation are complex. However, complex roots don't imply that (42) is satisfied. Consequently, the conditions for an oscillatory covariogram aren't quite equivalent with these for a spectral peak.

To illustrate the ability of low-order stochastic difference equations to generate "realistic" data, Figures 4a and 4b show simulations of first- and second-order stochastic difference equations, while Figure 4c shows the solution of the deterministic part of the same second-order difference equation with initial conditions $y_0=y_1=1$. Notice that even the first-order stochastic difference equation

$$y_t = .9y_{t-1} + \varepsilon_t,$$

ε_t a serially uncorrelated random term, appears to generate roughly alternating periods of boom and bust. This illustrates how stochastic difference equations can generate processes that "look like" they have business cycles even if their spectra don't have peaks on $(0,\pi)$ and even if their covariograms don't oscillate.

The Cross Spectrum

An alternative representation of the cross covariogram is provided by the cross spectrum. Recall that the cross covariance generating function between the jointly stationary processes y and x is defined by

$$g_{yx}(z) = \sum_{k=-\infty}^{\infty} c_{yx}(k)z^k.$$

If we evaluate $g_{yx}(z)$ at the value $z=e^{-iw}$, we have the cross spectrum

$$g_{yx}(e^{-iw}) = \sum_{k=-\infty}^{\infty} c_{yx}(k)e^{-iwk}.$$

Viewed as a function of angular frequency w, $g_{yx}(e^{-iw})$ is called the cross spectrum between y and x.

The cross spectrum is of course a cross-covariance generating function. Given an expression for $g_y(e^{-iw})$, it is possible to recover the cross covariances from the inversion formula

$$c_{yx}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} g_{yx}(e^{-iw})e^{iwk}dw.$$

The validity of this inversion formula can be checked by following calculations analogous to those used to verify the inversion formula for the spectrum.

Unlike the spectrum, the cross spectrum is in general a complex quantity at each frequency, this being a consequence of the fact that $c_{yx}(k)$ is in general not symmetric ($c_{yx}(k)$ does not in general equal $c_{yx}(-k)$). In place of the symmetry property, we have the readily verified property

$$(45) \qquad g_{xy}(e^{-iw}) = \overline{g_{yx}(e^{-iw})} = g_{yx}(e^{+iw})$$

where the bar denotes complex conjugation and

$$g_{xy}(e^{-iw}) = \sum_{k=-\infty}^{\infty} c_{xy}(k)e^{-iwk}$$

and $c_{xy}(k) = Ex_t y_{t-k}$. Notice that $c_{xy}(k) = c_{yx}(-k)$.

Suppose that the stationary stochastic process y_t is related to the stochastic processes x_t and ε_t by

$$(46) \qquad y_t = \sum_{j=-p}^{q} h_j x_{t-j} + \varepsilon_t$$

where $E\varepsilon_t = Ex_t = 0$, and $E\varepsilon_t x_{t-s} = 0$ for all s, an orthogonality condition that characterizes $\Sigma h_j x_{t-j}$ as the projection of y_t on the space $\{x_{t+p}, \ldots, x_{t-q}\}$. Then we have already seen that the spectrum of y satisfies

$$g_y(e^{-iw}) = |h(e^{-iw})|^2 \, g_x(e^{-iw}) + g_\varepsilon(e^{-iw})$$

where

$$h(e^{-iw}) = \sum_{j=-p}^{q} h_j e^{-iwj} \quad .$$

To find the cross spectrum between y and x, first use (46) to calculate the k^{th} lagged covariance as

$$Ey_t x_{t-k} = \sum_{j=-p}^{q} h_j E(x_{t-j} x_{t-k})$$

$$c_{yx}(k) = \sum_{j=-p}^{q} h_j c_x(k-j) \quad .$$

Thus the cross-covariogram between y and x is the convolution of the sequence $\{h_j\}$ with the sequence $c_x(j)$. From the convolution property we immediately have

$$g_{yx}(e^{-iw}) = h(e^{-iw}) \cdot g_x(e^{-iw})$$

since the Fourier transform of a convolution of two sequences is the product of the Fourier transforms of the two sequences. That is, taking Fourier transforms of each side (i.e., multiplying by e^{-iwk} and summing over k) gives

$$\sum_{k=-\infty}^{\infty} c_{yx}(k) e^{-iwk} = \sum_{j=-p}^{q} \sum_{k=-\infty}^{\infty} h_j c_x(k-j) e^{-iwk}$$

Noting that $e^{-iwk} = e^{-iw(k-j)} d^{-iwj}$, the above can be written as

$$g_{yx}(e^{-iw}) = \sum_{j=-p}^{q} h_j e^{-iwj} \sum_{k=-\infty}^{\infty} c_x(k-j) e^{-iw(k-j)}$$

or

(47) $$g_{yx}(e^{-iw}) = h(e^{-iw}) g_x(e^{-iw}) \quad .$$

Notice that the covariance between y and x can be recovered from the inversion formula

$$c_{yx}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} h(e^{-iw}) g_x(e^{-iw}) e^{iwk} dw$$

Further, notice that given $g_{yx}(e^{-iw})$ and $g_x(e^{-iw})$ the h_k's can be recovered from

$$h_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{g_{yx}(e^{-iw})}{g_x(e^{-iw})} e^{iwk} dw \quad .$$

Where estimators of $g_{yx}(e^{-iw})$ and $g_x(e^{-iw})$ are used in the above equation, the resulting estimator of the h_k's is known as Hannan's inefficient estimator.

Formula (22) can now be used to show that the spectrum reflects a decomposition of x_t into processes that are orthogonal across frequencies. Thus let

$$y_{1t} = B_1(L)x_t$$

$$y_{2t} = B_2(L)x_t$$

where $B_1(L)$ and $B_2(L)$ are chosen to satisfy

$$B_1(e^{-iw}) = \begin{cases} 1 & w\varepsilon[-b,-a]\cap[a,b] \\ 0 & w\not\varepsilon[-b,-a]\cap[a,b] \end{cases}$$

$$B_2(e^{-iw}) = \begin{cases} 1 & w\varepsilon[-d,-c]\cap[c,d] \\ 0 & w\not\varepsilon[-d,-c]\cap[c,d] \end{cases}$$

To find the individual distributed lag coefficients, equation (38) can be used. Equation (22) evaluated at $z=e^{-iw}$ implies

$$g_{y_1y_2}(e^{-iw}) = B_1(e^{-iw})B_2(e^{iw})g_x(e^{-iw})$$

If $[-b,-a]\cap[a,b]$ does not intersect with the set of frequencies $[-d,-c]\cap$ $[c,d]$, then $B_1(e^{-iw})B_2(e^{iw})=0$ for all w, so that $g_{y,y_2}(e^{-iw})=0$. This in turn implies that y_1 and y_2 are processes that are orthogonal (uncorrelated) at all lags, as can be verified directly from the inversion formula. In this sense the spectrum $g_x(e^{-iw})$ decomposes the variance of x into a set of mutually orthogonal processes across frequencies.

The cross spectrum is a complex quantity that is usually characterized by real numbers in various ways. One characterization is in terms of its real and imaginary parts

$$g_{yx}(e^{-iw}) = co(w) + iqu(w)$$

where co(w) is called the cospectrum and qu(w) is called the quadrature spectrum. A more usual representation is the polar one

$$g_{yx}(e^{-iw}) = r(w)e^{i\theta(w)} \tag{48}$$

where

$$r(w) = \sqrt{co(w)^2 + qu(w)^2}$$

$$\theta(w) = \tan^{-1}\left[\frac{qu(w)}{co(w)}\right] .$$

The phase statistic gives the lead of y over x at frequency w, while the "gain" r(w) tells how the amplitude in x is amplified in contributing to the amplitude of y at frequency w. Another interesting number is the coherence

$$coh(w) = \frac{|g_{yx}(e^{-iw})|^2}{g_x(e^{-iw})g_y(e^{-iw})} ,$$

which, being essentially the ratio of a covariance squared to the product of two variances, is analogous to an R^2 statistic. It indicates the proportion of the variance in one series at frequency w that is accounted for by variation in the other series.

Notice that from (47) and from the fact that the spectrum $g_x(e^{-iw})$ is real, the phase of the cross-spectrum equals the phase of $h(e^{-iw}) = \sum h_j e^{-iwj}$, which is the Fourier transform of the h_j's. That is, writing (47) and (48) we have

$$r(w)e^{i\theta(w)} = g_{yx}(e^{-iw}) = h(e^{-iw})g_x(e^{-iw})$$

or

$$h(e^{-iw}) = \frac{r(w)}{g_x(e^{-iw})} \cdot e^{i\theta(w)} ,$$

which shows that the phase of $g_{yx}(e^{-iw})$ equals the phase of $h(e^{-iw})$. For convenience, represent $h(e^{-iw})$ in the polar form

$$h(e^{-iw}) = s(w)e^{i\theta(w)}$$

where $s(w)=r(w)/g_x(e^{-iw})$.

The following provides a heuristic device for interpreting $\theta(w)$. Suppose we consider as an input into the system (46) an x series consisting of a pure cosine wave of frequency w:

$$x_t = 2\cos wt = e^{iwt} + e^{-iwt}$$

For this input path, suppressing the disturbance ε_t, (46) becomes

$$y_t = \sum h_j [e^{iw(t-j)} + e^{-iw(t-j)}]$$

$$= e^{iwt}\sum h_j e^{-iwj} + e^{-iwt}\sum h_j e^{+iwj}$$

But $\sum h_j e^{-iwj}=s(w)e^{i\theta(w)}$ and $\sum h_j e^{+iwj}$, being the complex conjugate of $\sum h_j e^{-iwj}$, equals $s(w)e^{-i\theta(w)}$. Therefore, we have

$$y_t = e^{iwt}s(w)e^{i\theta(w)} + e^{-iwt}s(w)e^{-i\theta(w)}$$

$$= s(w)[e^{i(wt+\theta(w))} + e^{-i(wt+\theta(w))}]$$

$$= s(w)2\cos(wt+\theta(w)).$$

Therefore, the response of (46) to an input in the form of a cosine wave of frequency w is a cosine wave at the same frequency with amplitude multiplied by s(w) and phase shifted by $\theta(w)$. The input cosine wave is at its peak at t=0, while the output is at its peak at $wt+\theta(w)=0$ or $t=-\frac{\theta(w)}{w}$ units of time. Thus, for $\theta(w)>0$, the output leads the input by $-\theta(w)/w$ units of time (where we adopt the usual convention that $\theta(w)$ is constrained to be between $-\pi$ and $+\pi$, a convention needed to make the arctangent function single-valued).

While useful, the preceding interpretation of the phase has to be used cautiously. The reason is that the stochastic difference equations that we have been studying generate random processes with spectral power distributed across a continuum of frequencies between $-\pi$ and $+\pi$. It is really only over a nonnegligible band of frequencies that there occurs a positive contribution to variance. Thus, for such processes there really don't occur input processes that are pure cosines, though this situation could be approached if the spectral density did display a very sharp peak at a given frequency. Processes with positive spectral power at a single given frequency do exist, and realizations of these processes do consist of (sums of) sine and cosine waves. But such processes aren't generated by the stochastic difference equations that we are studying.

It is interesting to note the following two facts about $h(e^{-iw})$. First, from the definition of $h(e^{-iw})$

$$h(e^{-iw}) = \sum_j h_j e^{-iwj},$$

we note that $h(e^{-iw})$ evaluated at w=0 is the sum of the lag weights, that is

$$h(e^{-i0}) = \sum h_j .$$

Notice that since

$$\sum h_j e^{-iwj} = \sum h_j \cos wj - i\sum h_j \sin wj$$

and that since sin 0=0, we have that

$$h(e^{-i0}) = s(0) = \sum h_j .$$

Since $h(e^{-iw})$ is real at zero frequency, the phase statistic $\theta(w)$ is zero at zero frequency:

$$(49) \qquad \theta(w) = \tan^{-1}\left[\frac{-\sum h_j \sin wj}{\sum h_j \cos wj}\right] .$$

$$\theta(0) = \tan^{-1}[0] = 0.$$

Next, it is possible to show that the derivative of the phase statistic with respect to w evaluated at w=0 equals minus the mean lag. Recall that

$$\frac{d}{dx} \tan^{-1} u = \frac{1}{1+u^2} \frac{du}{dx} .$$

Applying this to (43) gives

$$\theta'(w) = \frac{1}{1+\left[\frac{-\sum h_j \sin wj}{\sum h_j \cos wj}\right]^2} \left\{\frac{-\sum h_j \cos wj \sum h_j j\cos wj - \sum h_j j\sin wj \sum h_j \sin wj}{(\sum h_j \cos wj)^2}\right\}$$

Evaluating $\theta'(w)$ at w=0 gives

$$\theta'(0) = \frac{-\sum h_j j}{\sum h_j}$$

(Here we have used the facts that cos 0=1, sin 0=0.) The right side of this equation is minus the "mean lag" of the lag distribution formed by the h's, a statistic often reported in econometric studies involving estimates of distributed lags.

A Digression on Leading Indicators

For years, the National Bureau of Economic Research (NBER) has employed a number of heuristic techniques designed to isolate "leading indicators" of business cycle movements, presumably as an aid in the early recognition and prediction of cyclical movements. To translate into our vocabulary, essentially a good leading indicator displays a sizable phase lead at the low business cycle frequencies over some important "coincident" measures of the cycle like unemployment or GNP (as well as a large coherence with those coincident measures--so that the phase lead is not only large on average but is regular in its occurrence). While searching for leading indicators is perhaps an important thing to do by way of categorizing the data, it is important to recognize that a series y_t that displays a sizable phase lead over another series x_t at the most important business cycle frequencies does not necessarily help in predicting x_t any better than can be done by using past x's alone to predict x. We illustrate this fact with two examples.

First suppose we have the system governed by

$$(50) \qquad x_t = \lambda x_{t-1} + u_t \qquad\qquad |\lambda| < 1$$

$$y_t = h_0 x_t + h_1 x_{t-1} + \varepsilon_t$$

where $Eu_t = E\varepsilon_t = Eu_t\varepsilon_{t-s} = 0$ for all t and s, and where both u and x are serially uncorrelated. The cross spectrum between y and x is given by

$$g_{yx}(e^{-iw}) = (h_0+h_1e^{-iw})g_x(e^{-iw})$$

$$= (h_0+h_1\cos w-ih_1\sin w)g_x(e^{-iw})$$

$$= r(w)e^{i\theta(w)}g_x(e^{-iw})$$

where

$$r(w) = \sqrt{(h_0+h_1\cos w)^2+(h_1\sin w)^2}$$

$$\theta(w) = \tan^{-1}\left[\frac{-h_1\sin w}{h_0+h_1\cos w}\right] .$$

Now by suitably choosing h_0 and h_1, at a given frequency $\theta(w)$ can be set arbitrarily in the interval $(-\pi, \pi)$. This is in spite of the fact that the model (50) implies that y_t is of no use in terms of predicting x_t, for x_t is governed by a pure "autoregression," and depends only on itself lagged and the unpredictable random term u_t. Thus, even if y_t leads x_t at the low business cycle frequencies, it is of no use in predicting x_t.

To specialize this example somewhat, suppose we have

$$x_t = \lambda x_{t-1} + u_t$$

$$y_t = (x_t-x_{t-1}) + \varepsilon_t$$

where as before u and ε are mutually orthogonal (at all lags) white noise process. Calculating $h(e^{-iw})$, we have

$$h(e^{-iw}) = 1 - e^{-iw}$$

$$= e^{-i\frac{w}{2}}(e^{i\frac{w}{2}}-e^{-i\frac{w}{2}})$$

$$= e^{-i\frac{w}{2}} \cdot 2i \sin \frac{w}{2}$$

$$= 2e^{-i\frac{w}{2}} e^{i\frac{\pi}{2}} \sin \frac{w}{2}$$

$$= 2e^{i(\frac{\pi}{2} - \frac{w}{2})} \sin \frac{w}{2}$$

For low frequencies (i.e., those for which $\pi/2 > w/2$) the phase angle $\pi/2 - w/2 > 0$, implying that the output y leads x at the low frequency components. By making λ large enough, we can assure that these low frequencies account for most of the variation in x. Inspite of the fact that y leads x at these important low frequency components, y is of no use in predicting x once lagged x's are taken into account.

As our second example, consider the system

$$y_t = \sum_{j=-\infty}^{\infty} h_j x_{t-j} + \varepsilon_t$$

$$y_t = \lambda y_{t-1} + u_t$$

where we assume $E\varepsilon_t x_s = 0$ for all t,s, $Eu_t = 0$, and u_t is a white-noise stationary process. We further assume that

$$h_j = h_{-j} \qquad \text{for all } j \geq 1.$$

The cross spectrum between y and x is calculated to be

$$g_{yx}(e^{-iw}) = \{h_0 + h_1(e^{iw} + e^{-iw}) + h_2(e^{2iw} + e^{-2iw}) + \ldots\} \, g_x(e^{-iw})$$

$$= (h_0 + 2\sum_{j=1}^{n} h_j \cos wj) g_x(e^{-iw})$$

which is <u>real</u> for all w. Therefore, the phase shift $\theta(w) = 0$ for all w, so that y and x are perfectly in phase at all frequencies. Despite this, by using a theorem due to Sims (see below) it is possible to show that even given the past of x, past y does help predict present and future x's. This is a consequence of the lag distribution of h_j's being two-sided and of Sims's theorem 2, which we will describe in detail shortly.

Taken together, these two examples illustrate the fact that displaying a phase lead is neither a necessary nor a sufficient condition for one series to be of use in predicting another.

Analysis of Some Filters: The Slutsky Effect and Kuznets' Transformations

Relation (36) can be used to show the famous "Slutsky effect." Slutsky considered the effects of starting with a white noise ε_t, taking a 2 period moving sum n times, and then taking first differences m times. That is, Slutsky considered forming the series

$$Z_t = (1+L)(1+L) \ldots (1+L)\varepsilon_t = (1+L)^n \varepsilon_t$$

and

$$y_t = (1-L)(1-L) \ldots (1-L)Z_t = (1-L)^m Z_t$$

(51) $$y_t = (1+L)^n (1-L)^m \varepsilon_t .$$

Applying (36) to (51) we have

$$g_y(e^{-iw}) = (1+e^{iw})^n (1+e^{-iw})^n (1-e^{iw})^m (1-e^{-iw})^m \sigma_\varepsilon^2$$

$$= [(1+e^{iw})(1+e^{-iw})]^n [(1-e^{-iw})(1-e^{iw})]^m \sigma_\varepsilon^2$$

$$= [(2+(e^{iw}+e^{-iw})]^n [(2-(e^{iw}+e^{-iw})]^m \sigma_\varepsilon^2$$

(52) $$g_y(e^{-iw}) = \sigma_\varepsilon^2 2^n[1+\cos w]^n\ 2^m[1-\cos w]^m .$$

Consider first the special case where m=n. Then (52) becomes

$$g_y(e^{-iw}) = \sigma_\varepsilon^2 4^n[1-\cos^2 w]^n$$

(53) $$= \sigma_\varepsilon^2 4^n[\sin^2 w]^n .$$

On $[0,\pi]$, the spectrum of y has a peak at $w=\pi/2$, since there sin w=1. Notice that since $\sin w\leq 1$, (53) implies that as n becomes large, the peak in the spectrum of y at /2 becomes sharp. In the limit, as $n\to\infty$, the spectrum of y becomes a "spike" at $\pi/2$, which means that y behaves like a cosine of angular frequency $\pi/2$.

Similar behavior results for fixed m/n as n becomes large where $m\neq n$. Consider (52) and set $dg_y(e^{-iw})/dw$ equal to zero in order to locate the peak in the spectrum:

$$\frac{dg_y}{dw} = \sigma_\varepsilon^2 2^{m+n}\{n[1-\cos w]^m[1+\cos w]^{n-1}(-\sin w)$$

$$+ m(1-\cos w)^{m-1}(\sin w)[1+\cos w]^n\}$$

$$= \sigma_\varepsilon^2 2^{m+n}\sin w\{(1-\cos w)^{m-1}(1+\cos w)^{n-1}\cdot$$

$$[m(1+\cos w)-n(1-\cos w)]\} .$$

This expression can equal zero on $(0,\pi)$ only if the expression in brackets equals zero:

$$m(1+\cos w)-n(1-\cos w) = 0$$

which implies

$$\cos w = \frac{1 - \frac{m}{n}}{1 + \frac{m}{n}},$$

or

$$w = \cos^{-1}\left(\frac{1-m/n}{1+m/n}\right)$$

which tells us the frequency at which the spectrum of y attains a peak. For fixed m/n, the spectrum of y approaches a spike as $n \to \infty$. This means that as $n \to \infty$, y tends to behave more and more like a cosine of angular frequency $\cos^{-1}((1-m/n)/(1+m/n))$.

What Slutsky showed, then, is that by successively summing and then successively differencing a serially uncorrelated or "white noise" process ε_t, a series with "cycles" is obtained.

Another use of (36) is in the analysis of transformations that have been applied to data. An example is Howrey's analysis of the transformations used by Kuznets. Data constructed by Kuznets have been inspected to verify the existence of "long swings," long cycles in economic activity of around twenty years. Before analysis, however, Kuznets subjected the data to two transformations. First, he took a five-year moving average:

$$Z_t = \frac{1}{5}[L^{-2}+L^{-1}+1+L+L^2]X_t \equiv A(L)X_t.$$

Then he took the centered first difference of the (nonoverlapping) five-year moving average:

$$y_t = Z_{t+5} - Z_{t-5} = [L^{-5}-L^5]Z_t = B(L)Z_t.$$

So we have that the y's are related to the X's by

$$y_t = \frac{1}{5}[L^{-5}-L^{5}][L^{-2}+L^{-1}+1+L+L^{2}]X_t$$

$$= A(L)B(L)X_t.$$

The spectrum of y is related to the spectrum of X by

$$g_y(e^{-iw}) = A(e^{-iw})A(e^{iw})B(e^{-iw})B(e^{iw})g_x(e^{-iw}). \tag{54}$$

We have

$$A(e^{-iw}) = \frac{1}{5}\sum_{j=-2}^{2} e^{-iwj} = \frac{1}{5}\frac{(e^{iw2}-e^{-iw3})}{(1-e^{-iw})}.$$

Thus,

$$A(e^{-iw})A(e^{iw}) = \frac{(\frac{1}{5})^2(e^{iw2}-e^{-iw3})(e^{-iw2}-e^{iw3})}{(1-e^{-iw})(1-e^{iw})}$$

$$= \frac{(\frac{1}{5})^2(2-(e^{iw5}+e^{-iw5}))}{(2-(e^{iw}+e^{-iw}))}$$

$$= \frac{(\frac{1}{5})^2 2(1-\cos 5w)}{2(1-\cos w)} = \frac{(\frac{1}{5})^2(1-\cos 5w)}{(1-\cos w)}.$$

Next, we have

$$B(e^{-iw}) = (e^{+iw5}-e^{-iw5})$$

so that

$$B(e^{-iw})B(e^{iw}) = (e^{iw5}-e^{-iw5})(e^{-iw5}-e^{iw5})$$

$$= (2-(e^{iw10}+e^{-iw10})) = 2(1-\cos 10w).$$

So it follows from (49) that

$$g_y(e^{-iw}) = [\frac{(\frac{1}{5})^2(1-\cos 5w)2}{(1-\cos w)} (1-\cos 10w)]g_x(e^{-iw})$$

$$= G(w)g_x(e^{-iw}).$$

where $G(w)=2[(\frac{1}{5})^2(1-\cos 5w)(1-\cos 10w)/(1-\cos w)]$. The term G(w) is graphed in Figure 5. It has zeroes at values where cos 5w=1 and where cos 10w=1. The first condition occurs on $[0,\pi]$ where

$$5w = 0,\ 2\pi,\ 4\pi,$$

or

$$w = 0,\ \frac{2}{5}\pi,\ \frac{4}{5}\pi.$$

The condition cos 10w=1 on $[0,\pi]$ where

$$10w = 0,\ 2\pi,\ 4\pi,\ 6\pi,\ 8\pi,\ 10\pi$$

or

$$w = 0,\ \frac{1}{5}\pi,\ \frac{2}{5}\pi,\ \frac{4}{5}\pi,\ \text{and}\ \pi.$$

So G(w) has zeroes at $w=0$, $\pi/5$, $2/5\pi$, $3\pi/5$, $4\pi/5$, and π.

From the graph of G(w), it follows that even if X_t is a white noise, a y series generated by applying Kuznets' transformations will have a large peak at a low frequency, and hence will seem to be characterized by "long swings." These long swings are clearly a statistical artifact; that is, they are something induced in the data by the transformation applied and not really a characteristic of the economic system.

With annual data, the biggest peak in Figure 5 corresponds to a cycle of about 20 1/4 years which is close to the 20-year cycle found by Kuznets. Howrey's observations naturally raise questions about the authenticity of the long swings identified by studying the data used by Kuznets.

A Small Kit of $h(e^{-iw})$'s

In order to provide some feel for the effects of various commonly used filters Figure 6 reports the amplitude and phase of $h(e^{-iw})$ for various h(L) lag distributions.

We have already calculated that for h(L)=1-L,

$$h(e^{-iw}) = 2e^{i(\frac{\pi}{2} - \frac{w}{2})} \sin\frac{w}{2},$$

as the graphs confirm.

For h(L)=1+L it is straightforward to calculate

$$h(e^{-iw}) = 1 + e^{-iw} = e^{-i\frac{w}{2}}(e^{+i\frac{w}{2}}+e^{-i\frac{w}{2}})$$

$$= 2e^{-i\frac{w}{2}}\cos\frac{w}{2}$$

which again agrees with our graphs.

Notice that for $h(L) = (1-t_1L-t_2L^2)^{-1}$, we have chosen (t_1, t_2) in the regions of peaked spectra of our figure (2). Notice that as required, $h(e^{-iw})$ is characterized by peaks. (See Figure (2)).

Alternative Definitions of the Business Cycle

We have already encountered two definitions of a cycle in a single series that is governed by a stochastic difference equation. According to the first definition, a variable possesses a cycle of a given frequency if its covariogram displays damped oscillations of that frequency, which is equivalent with the condition that the nonstochastic part of the difference equation has a pair of complex roots with argument (θ in the polar form of the root $re^{i\theta}$) equal to the frequency in question. A single series is said to contain a <u>business cycle</u> if the cycle in question has periodicity of from about two to four years (NBER minor cycles) or about eight years (NBER major cycles).

A second definition of a cycle in a single series is the occurrence of a peak in the spectral density of a series. As we have seen, this definition is not equivalent with the previous one, but usually leads to a definition of the cycle close to the first one.

It is probably correct, however, that neither one of these definitions is what underlies the concept of the business cycle that most experts have in mind. In fact, most economic aggregates have spectral densities that do not display pronounced peaks at the range of frequencies associated with the business cycle. The peaks that do occur in this band of frequencies tend to be wide and of modest height. The dominant feature of the spectrum of most economic time series is that it generally decreases drastically as frequency increases, with most of the power in the low frequency, high periodicity bands. This shape was dubbed by Granger the "typical spectral shape" of an economic variable, and is illustrated by the spectral density of the monthly average call rate over the period 1890-1913, which is shown in Figure 7. The generally downward sweeping spectrum is characteristic of a covariogram that is dominated by high positive, low-order serial correlation. (The call rate spectrum displays a second feature that is often possessed by spectra of economic time series: peaks at the seasonal frequencies of 12, 6, 4, 3, 2.4, and 2 months.) As mentioned earlier, the fact that a spectrum doesn't display a peak at the business cycle frequencies should not be taken to mean that the series didn't experience any fluctuations associated with the business cycle. On the contrary, as on Figure 4a indicated, a series could very well seem to move in sympathy with general business conditions say as identified by the NBER and yet have no spectral peak on the open internal $(0,\pi)$. This example cautions the reader against interpreting the lack of a peak in the spectrum at the business cycle frequencies as indicating the absence of any business cycle in the series.

What the preceding example does indicate is that our two preceding possible definitions of the business cycle are deficient. The following definition seems to capture what experts refer to as the business cycle: the business cycle is the phenomenon of a number of important economic aggregates (such as GNP, unemployment, and layoffs) being characterized by high pairwise coherences at the low business cycle frequencies, the same frequencies at which most aggregates have most of their spectral power if they have "typical" spectral shapes. This definition captures the notion of the business cycle as being a condition symptomizing the common movements of a set of aggregates.

Representation Theory

So far we have generally started with a white noise ε_t as a building block and considered constructing a stochastic process x_t via a transformation

$$x_t = B(L)\varepsilon_t \quad .$$

In this section we reverse this procedure and start out by assuming that we have a covariance stationary process x_t with covariogram $c(\tau)$. We then show that associated with every such process $\{x_t\}$ is a white noise process $\{\varepsilon_t\}$ that is its fundamental building block. One purpose of this construction is to convey the sense in which the models we have been studying are quite general ones for covariance stationary processes.

Suppose that we have a covariance stationary stochastic process x_t with covariogram $c(\tau)$ and mean zero. We think of forming a sequence of linear least squares projections of x_t against a sequence of expanding sets of past x's, $\{x_{t-1}, x_{t-2}, \ldots, x_{t-n}\}$:

$$\hat{x}_t^n = \sum_{i=1}^{n} a_i^n x_{t-i} = P[x_t | x_{t-1}, \ldots, x_{t-n}]$$

or

$$x_t = \hat{x}_t^n + \varepsilon_t^n$$

where $E\varepsilon_t^n x_{t-i} = 0$ for $i=1,\ldots,n$ by the orthogonality principle. These orthogonality conditions uniquely determine the projection $\hat{x}_t^n = \sum_{i=1}^{n} a_i^n x_{t-i}$. The population covariogram $c(\tau)$ contains all of the information necessary to calculate the a_i^n's from the least squares normal equations.*

As n is increased toward infinity, it is possible to show that the sequence of projections $\{\hat{x}_t^n\}$ converge to a random variable $\hat{x}_t$ in the "mean square" sense that**

$$\lim_{n\to\infty} E\,(\hat{x}_t - \hat{x}_t^n)^2 = 0 .$$

This means that for any $\delta>0$, we can find an $N(\delta)$ such that

$$E(\hat{x}_t - \hat{x}_t^m)^2 < \delta$$

for all $m>N(\delta)$, so that in the mean square sense, we can approximate arbitrarily well the projection in the space spanned by the infinite set of lagged x's with the projection of x_t on a suitable finite set of lagged x's.† We write the projection of x_t on the space spanned by the infinite set $(x_{t-1}, x_{t-2}, \ldots)$ as

$$\hat{x}_t = P[x_t | x_{t-1}, x_{t-2}, \ldots]$$

---

* The a_i^n's will be unique only if there are no linear dependencies across the x_{t-i}'s. The projection of x_t on the space spanned by $\{x_{t-1},\ldots,x_{t-n}\}$ is unique even without that condition.

** It is not necessarily true that the sequence of (a_i^n)'s settles down nicely as $n\to\infty$, only that successive $\hat{x}_t^n$'s get closer to each other and to $\hat{x}_t$ as $n\to\infty$.

† For a proof, see T. W. Anderson, The Statistical Analysis of Time Series, Wiley, p. 419.

and have the decomposition of x_t as

(55) $$x_t = P[x_t | x_{t-1}, x_{t-2}, \ldots] + \varepsilon_t$$

where ε_t is a least squares residual that obeys the orthogonality condition $E\varepsilon_t x_{t-i} = 0$ for all $i \geq 1$. In mean square, ε_t is the limit as $n \to \infty$ of ε_t^n, i.e. $\lim_{n\to\infty} E(\varepsilon_t - \varepsilon_t^n)^2 = 0$.

We can now state an important decomposition theorem due to Wold.

<u>Theorem</u>: Let $\{x_t\}$ be any covariance stationary stochastic process with $Ex_t = 0$. Then it can be written as

$$x_t = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j} + \eta_t$$

where $d_0 = 1$, where $\sum_{j=0}^{\infty} d_j^2 < \infty$, $E\varepsilon_t^2 = \sigma^2 \geq 0$, $E\varepsilon_t \varepsilon_s = 0$ for $t \neq s$ (so that $\{\varepsilon_t\}$ is serially uncorrelated), $E\varepsilon_t = 0$ and $E\eta_t \varepsilon_s = 0$ for all t and s (so that $\{\varepsilon\}$ and $\{\eta\}$ are processes that are orthogonal at all lags); and $\{\eta_t\}$ is a process that can be predicted arbitrarily well by a linear function of only past values of x_t, i.e., η_t is linearly deterministic. Furthermore, $\varepsilon_t = x_t - P[x_t | x_{t-1}, x_{t-0}, \ldots]$.

<u>Proof</u>: We let ε_t be the same ε_t as appears in (55), so that

$$\varepsilon_t = x_t - P[x_t | x_{t-1}, x_{t-2}, \ldots]$$

So ε_t is the error or "innovation" in predicting x_t from its own past. Now ε_t is orthogonal to $\{x_{t-1}, x_{t-2}, \ldots\}$, by the orthogonality principle. But ε_{t-s} is a linear combination of past x's:

$$\varepsilon_{t-s} = x_{t-s} - P[x_{t-s}|x_{t-s-1},\ldots] \quad .$$

Therefore $E\varepsilon_t\varepsilon_{t-s} = 0$ for all t and s. So we have proved that $\{\varepsilon_t\}$ is a serially uncorrelated process.

Now think of projecting x_t against a sequence of sets spanned by $(\varepsilon_t,\varepsilon_{t-1},\ldots,\varepsilon_{t-m})$ for successively larger m's. The typical projection of x_t on such a set is

$$\hat{x}_t^m = \sum_{j=0}^{m} d_j \varepsilon_{t-j}$$

where, since the ε_{t-j}'s are mutually orthogonal, the d_j's are given by

$$d_j = \frac{Ex_t\varepsilon_{t-j}}{\sigma^2}$$

$$\sigma^2 = E\varepsilon_t^2 \quad .$$

Notice that since $\varepsilon_t = x_t - P[x_t|x_{t-1},x_{t-2},\ldots]$ and since $E\varepsilon_t x_{t-i} = 0$ for all $i\geq 1$, we have $E\varepsilon_t^2 = Ex_t\varepsilon_t$. Thus, we have $d_0 = Ex_t\varepsilon_t/E\varepsilon_t^2 = 1$. Since the ε's are orthogonal, the d_j's don't depend on m. Now calculate the variance of the prediction error, which is

$$E(x_t - \sum_{j=0}^{m} d_j\varepsilon_{t-j})^2$$

$$= Ex_t^2 - 2E\sum_{j=0}^{m} d_j Ex_t\varepsilon_{t-j} + E\left[\sum_{j=0}^{m} d_j^2\varepsilon_{t-j}^2\right]$$

$$= Ex_t^2 - 2\sigma^2\sum_{j=0}^{m}\left(\frac{Ex_t\varepsilon_{t-j}}{\sigma^2}\right)^2 + \sigma^2\sum_{j=0}^{m}\left(\frac{Ex_t\varepsilon_{t-j}}{\sigma^2}\right)^2$$

$$= Ex_t^2 - \sigma^2\sum_{j=0}^{m} d_j^2 \geq 0 \;,$$

where the last inequality follows because the variance of the prediction error cannot be negative. Since $Ex_t^2 < \infty$, from the last inequality it follows that for all m

$$\sigma^2 \sum_{j=0}^{m} d_j^2 < Ex_t^2$$

so that $\sum_{j=0}^{\infty} d_j^2 < \infty$. It follows that $\sum_{j=0}^{\infty} d_j \varepsilon_{t-j}$ is well defined, i.e. it converges in the mean square sense.*

Now define the process η_t by

$$\eta_t = x_t - \sum_{j=0}^{\infty} d_j \varepsilon_{t-j} \quad .$$

---

* That is, the sequence of $\sum_{j=0}^{m} d_j \varepsilon_{t-j}$'s is a Cauchy sequence. In particular, for n>m

$$E\left(\sum_{j=0}^{m} d_j \varepsilon_{t-j} - \sum_{j=0}^{n} d_j \varepsilon_{t-j} \right)^2$$

$$= E\left(\sum_{j=m+1}^{n} d_j^2 \varepsilon_{t+j}^2 \right)$$

$$= \sigma^2 \sum_{j=n+1}^{n} d_j^2 \quad .$$

Since $\sum_{j=0}^{\infty} d_j^2 < \infty$, it follows that we can choose an m big enough to drive

$\sigma^2 \sum_{j=m+1}^{\infty} d_j^2$ arbitrarily close to zero.

Notice that for s≤t we have

$$E\eta_t\varepsilon_s = Ex_t\varepsilon_s - E\sum_{j=0}^{\infty} d_j\varepsilon_s\varepsilon_{t-j}$$

$$= Ex_t\varepsilon_s - d_{t-s}E\varepsilon_s^2$$

$$= Ex_t\varepsilon_s - Ex_t\varepsilon_s = 0 \quad .$$

In addition $E\eta_t\varepsilon_s = 0$ for all s>t because ε_s is orthogonal to all x's dated earlier than s and by construction η_t is in the space spanned by x's dated t and earlier. Thus $\{\eta_t\}$ is orthogonal to $\{\varepsilon_t\}$ at all lags and leads. That is, the entire $\{\varepsilon\}$ process is orthogonal to the entire $\{\eta\}$ process.

Because η_t is orthogonal to ε_t, η_t must lie in the space spanned by $\{x_{t-1}, x_{t-2}, \ldots\}$ since square summable* linear combinations of $\{x_{t-1}, x_{t-2}, \ldots\}$ form the space of <u>all</u> random variables orthogonal to ε_t.** This implies that η_t can be predicted perfectly from lagged x's. More precisely, project $\eta_t = x_t - \sum_{j=0}^{\infty} d_j\varepsilon_{t-j}$ against $\{x_{t-1}, x_{t-2}, \ldots\}$ to get

$$P[\eta_t|x_{t-1},\ldots] = P[x_t|x_{t-1},\ldots] - \sum_{j=1}^{\infty} d_j\varepsilon_{t-j}$$

since $P[\varepsilon_t|x_{t-1},\ldots] = 0$ and since $P[\varepsilon_{t-k}|x_{t-1},\ldots] = \varepsilon_{t-k}$ for k≥1. Subtracting the above equation from the definition of η_t gives

$$\eta_t - P[\eta_t|x_{t-1},\ldots] = (x_t - P[x_t|x_{t-1},\ldots]) - d_0\varepsilon_t = 0$$

---

* Those linear combinations $\sum_{j=1}^{\infty} f_j x_{t-j}$ for which $\sum_{j=1}^{\infty} f_j^2 < \infty$, so that the variance of the sum is finite.

** This is an implication of the orthogonality principle. See T.W. Anderson, p. .

since the one-step-ahead prediction error for x_t is $d_0\varepsilon_t$. Thus, $\eta_t = P[\eta_t|x_{t-1},\ldots]$ so that η_t can be predicted arbitrarily well (in the mean squared error sense) from past x's alone. More generally, we have

$$P[\eta_t|x_{t-k},x_{t-k-1},\ldots] = P[x_t|x_{t-k},\ldots] + \sum_{j=k}^{\infty} d_j\varepsilon_{t-j} \quad .$$

Subtracting this from the definition of η_t gives

$$\eta_t - P[\eta_t|x_{t-k},\ldots] = (x_t - P[x_t|x_{t-k},\ldots]) - \sum_{j=0}^{k-1} d_j\varepsilon_{t-j} = 0 \quad ,$$

since $\sum_{j=0}^{k-1} d_j\varepsilon_{t-j}$ is the k step ahead prediction error in predicting x_t from its own past. Thus, we have proved that η_t is (linearly) deterministic in the sense that it can be predicted arbitrarily well (in the mean squared error sense) arbitrarily far into the future from past x's only. This completes the proof of Wold's theorem.

The η_t process is termed the (linearly) deterministic part of x_t while $\sum_{j=0}^{\infty} d_j\varepsilon_{t-j}$ is termed the (linearly) indeterministic part. The reason for the adverb "linearly" is that the decomposition has been obtained by using linear projections.

Wold's theorem is important for us because it provides an explanation of the sense in which stochastic difference equations provide a general model for the indeterministic part of any univariate stationary stochastic process, and also the sense in which there exists a white noise process ε_t that is the building block for the indeterministic part of x_t. Not surprisingly, the construction of the theorem can be extended to multivariate stochastic processes for which a corresponding orthogonal decomposition exists in which the deterministic and indeterministic parts are vectors.

As a particular example of a process that conforms to the representation given in Wold's decomposition theorem, consider the process

$$x_t = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j} + \sum_{k=1}^{n} (a_i \cos \lambda_i t + b_i \sin \lambda_i t)$$

where ε_t is covariance stationary, serially uncorrelated process with mean zero and variance σ_ε^2, $\sum_{j=0}^{\infty} d_j^2 < \infty$, a_i and b_i are random variables orthogonal to the entire ε process and satisfying $Ea_i = Eb_i = Ea_i b_i = 0$ for all i; $Ea_i a_j = Eb_i b_j = 0$ for all i,j; and $Ea_i^2 = Eb_i^2 = \sigma_i^2$; the λ_i's are fixed numbers in the interval $[-\pi, \pi]$. The process $\sum_{i=1}^{n} (a_i \cos \lambda_i t + b_i \sin \lambda_i t)$ is deterministic, is orthogonal to the process $\sum d_j \varepsilon_{t-j}$ at all lags, and is easily deduced* to have covariogram given by

---

* For example, let

$$x(t) = a \cos \lambda t + b \sin \lambda t$$

where $Ea = Eab = Eb = 0$, $Ea^2 = Eb^2 = \sigma^2$. Then

$$\begin{aligned} Ex(t_1)x(t_2) &= E\{a^2 \cos \lambda_1 t \cos \lambda t_2 + \\ &\quad ab(\cos \lambda t_2 \cos \lambda t_1 + \sin \lambda t_2 \sin \lambda t_1) \\ &\quad + b^2 \sin \lambda t_1 \sin \lambda t_2\} \\ &= \sigma^2 \{\cos \lambda t_1 \cos \lambda t_2 + \sin \lambda t_1 \sin \lambda t_2\} \end{aligned}$$

Since $\cos(\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta$, we have

$$Ex(t_1)x(t_2) = \sigma^2 \cos \lambda (t_1 - t_2)$$

or $$Ex(t)x(t-T) = {}^2 \cos \lambda T \quad .$$

These calculations can easily be extended to prove the assertion made in the text.

$\sum_{i=1}^{n} \sigma_i^2 \cos \lambda_i \tau$. As we have seen, the covariogram of $\sum_{j=0}^{\infty} d_j \varepsilon_{t-j}$ has generating function $\sigma_\varepsilon^2 d(z) d(z^{-1})$. The spectral density of the deterministic part turns out not to be well defined as an ordinary function. This can be seen by noting that the ordinary Fourier transform of the covariogram $\sigma^2 \cos \lambda_i \tau$ is

$$\sigma^2 \sum_{\tau=-\infty}^{\infty} \cos \lambda \tau e^{-iw\tau} = \sigma^2 \sum_{\tau=-\infty}^{\infty} \left(\frac{e^{i\lambda\tau} + e^{-i\lambda\tau}}{2} \right) e^{-iw\tau}$$

$$= \sigma^2 \sum_{\tau=-\infty}^{\infty} \left\{ \frac{e^{i(\lambda-w)\tau} + e^{-i(\lambda+w)\tau}}{2} \right\}$$

Notice that the first term can be written

$$\sum_{\tau=-\infty}^{\infty} e^{i(\lambda-w)\tau} = 1 + \sum_{\tau=1}^{\infty} \left\{ e^{i(\lambda-w)\tau} + e^{-i(\lambda-w)\tau} \right\}$$

$$= 1 + 2 \sum_{\tau=1}^{\infty} \cos(\lambda-w)\tau \quad .$$

The series $\sum_{\tau=1}^{\infty} \cos(\lambda-w)\tau$ is <u>not</u> a convergent series, so that the spectrum of the deterministic part of our process is not well defined by the usual Fourier transformation.

However, it happens that there is a sense in which the spectrum of the deterministic part does exist, namely, in the sense of a generalized function or "distribution." In particular, let $\delta(w)$ be the delta generalized function which has "infinite mass" at w=0 and is zero everywhere else. That is, $\delta(w)$ is defined by

$$\int_{-\infty}^{\infty} \delta(w) g(w) dw = g(0)$$

where g(w) is any ordinary "test function" that is continuous at zero. Then the spectral density of a process with covariogram $\sigma^2 \cos \lambda\tau$ is defined as

$$f(w) = 2\pi \left\{ \frac{\sigma^2}{2} \delta(w-\lambda) + \frac{\sigma^2}{2} \delta(w+\lambda) \right\} .$$

With the spectral density so defined, notice that the inversion formula holds, i.e.

$$c(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(w) e^{iw\tau} dw$$

$$= \frac{\sigma^2}{2} \left\{ \int_{-\infty}^{\infty} \delta(w-\lambda) e^{iw\tau} dw + \int_{-\infty}^{\infty} \delta(w+\lambda) e^{+iw\tau} dw \right\}$$

$$= \sigma^2 \left\{ \frac{e^{i\lambda\tau} + e^{-i\lambda\tau}}{2} \right\}$$

$$= \sigma^2 \cos \lambda\tau \quad .$$

Then the spectral density of the deterministic part of our process is

$$2\pi \sum_{i=1}^{n} \sigma_i^2 \left\{ \frac{\delta(w-\lambda_i)}{2} + \frac{\delta(w+\lambda_i)}{2} \right\} ,$$

so that the spectral density function of the deterministic part is zero except for the singular points $w=\pm\lambda_i$, $i=1,\ldots,n$, at which the spectrum has mass $\sigma_i^2/2$. The spectral density thus has "spikes" at the points $w=\pm\lambda_i$.*

---

*There are essentially two ways in which a process can be deterministic. One is if its spectral density consists entirely of a number of "spikes" or delta functions. A second way is if its spectral density, even though having no spikes, is zero on some interval of w's of positive length, or is "too close" to zero over such an interval.

Linear Least Squares Prediction

It is common in economics to assume that x_t is purely (linearly) indeterministic, which means that $\eta_t = 0$ for all t, or else that η_t has been removed.* Wold's theorem says that <u>any</u> indeterministic covariance stationary stochastic process x_t has the <u>moving average representation</u>

$$x_t = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j}$$

or

$$(56) \qquad x_t = d(L)\varepsilon_t \quad , \qquad d(L) = \sum_{j=0}^{\infty} d_j L^j$$

where $\{\varepsilon_t\}$ is the sequence of one-step ahead linear least squares forecasting errors (innovations) in predicting x_t as a linear function of $\{x_{t-1}, x_{t-2}, \ldots\}$. (As we have seen, it is natural to normalize d(L) so that $d_0=1$, in which case $\sigma^2 = E\varepsilon_t^{\ 2}$ is the variance of the one-step ahead prediction error.)

Now suppose that d(L) has an inverse that is one-sided in nonnegative powers of L. Where $d(L) = \sum_{j=0}^{n} d_j L^j$, a necessary and sufficient condition for d(L) to have such a one-sided inverse is that the roots μ of

$$\sum_{j=0}^{n} d_j \mu^j = 0$$

all lie outside the unit circle, i.e. all have absolute values exceeding

---

* For example, by suitable detrending and seasonal adjustment.

unity. An inverse $a(L) \equiv d(L)^{-1}$ of $d(L)$ satisfies $a(L)d(L) = d(L)a(L) = I$ where I is the identity lag operator $I = 1 + 0L + 0L^2 + \dots$. Operating on both sides of (56) with $a(L) = d(L)^{-1}$ gives

$$(57) \qquad a(L)x_t = \varepsilon_t \quad , \quad a(L) = a_0 - \sum_{j=1}^{\infty} a_j L^j$$

or

$$a_0 x_t = +a_1 x_{t-1} + a_2 x_{t-2} + \dots + \varepsilon_t \quad .$$

Since d_0 is unity, it turns out that a_0 is unity also. Equation (57) is termed the autoregressive representation for x_t. While every linearly indeterministic covariance stationary process has a moving average representation, not all of them have an autoregressive representation. Still, those that do have both a moving average and an autoregressive representation constitute a very wide class, and we shall henceforth assume that we're dealing with a member of this class.*

We now derive some formulas due to Wiener and Kolmogorov for linear least squares predictors. Let $P_{t-j}x_t$ be the linear least squares projection of x_t on the space spanned by $\{x_{t-j}, x_{t-j-1}, \dots\}$; i.e.

$$P_{t-j}x_t \equiv P[x_t | x_{t-j}, x_{t-j-1}, \dots] \quad .$$

---

* We remarked earlier that in general the sequence of (a_j^n's) in

$$P[x_t | x_{t-1}, \dots, x_{t-n}] = \sum_{j=1}^{n} a_j^n x_{t-j}$$

does not converge as $n \to \infty$. However, under the roots condition given in the text, the a_j^n's do converge. In particular, they converge to the a_j's of equation (57), so that $\lim_{n\to\infty} a_j^n = a_j$ for all $j=1,2,\dots$.

Now project both sides of (56) against $\{x_{t-1}, x_{t-2}, \ldots\}$ to get

$$P_{t-1}x_t = \sum_{j=0}^{\infty} d_j P_{t-1}\varepsilon_{t-j}$$

$$= \sum_{j=1}^{\infty} d_j \varepsilon_{t-j}$$

which follows since $P_{t-1}\varepsilon_t = 0$, because ε_t is orthogonal to lagged x's; and since $P_{t-1}\varepsilon_{t-j} = \varepsilon_{t-j}$ for all $j \geq 1$, because ε_{t-j} is in the space spanned by $\{x_{t-1}, x_{t-2}, \ldots\}$. We write the above equation as

$$P_{t-1}x_t = \left(\sum_{j=1}^{\infty} d_j L^j\right)\varepsilon_{t-1}$$

$$P_{t-1}x_t = \left(\frac{d(L)}{L}\right)_+ \varepsilon_{t-1}$$

where $(\quad)_+$ means "ignore negative powers of L." Now assuming that x_t has an autoregressive representation, we can write $\varepsilon_{t-1} = a(L)x_{t-1}$ $= d(L)^{-1}x_{t-1}$. Substituting this into the above equation gives

$$P_{t-1}x_t = \left(\frac{d(L)}{L}\right)_+ \frac{1}{d(L)} x_{t-1}\ , \tag{58}$$

which is a compact formula for the one-step ahead linear least squares forecast of x_t based on its own past.

To get a formula for the general k-step ahead linear least squares forecast, project both sides of (56) against $\{x_{t-k}, x_{t-k-1}, \ldots\}$ to get

$$P_{t-k}x_t = \sum_{j=k}^{\infty} d_j \varepsilon_{t-j} = \left(\frac{d(L)}{L^k}\right)_+ \varepsilon_{t-k}$$

$$P_{t-k}x_t = \left(\frac{d(L)}{L^k}\right)_+ \frac{1}{d(L)} x_{t-k} \tag{59}$$

which generalizes formula (58).

Some Examples

First-order Markov

Consider the first-order autoregressive process

$(1-\lambda L)x_t = \varepsilon_t$, ε_t white noise, $|\lambda|<1$.

$$x_t = \left(\frac{1}{1-\lambda L}\right)\varepsilon_t \quad .$$

We have

$$\begin{aligned} P_{t-1}x_t &= [L^{-1}(1+\lambda L + \lambda^2 L^2 + \ldots)]_+ \; (1-\lambda L)x_{t-1} \\ &= (\lambda + \lambda^2 L + \ldots)(1 - \lambda L)x_{t-1} \\ &= \left(\frac{\lambda}{1-\lambda L}\right)(1-\lambda L)x_{t-1} \\ &= \lambda x_{t-1} \end{aligned}$$

More generally,

$$\begin{aligned} P_{t-k}x_t &= [L^{-k}(1 + \lambda L + \ldots)]_+ \; (1-\lambda L)x_{t-1} \\ &= \lambda^k x_{t-k} \quad . \end{aligned}$$

Thus we have

$$P_t x_{t+k} = \lambda^k x_t \quad .$$

First order moving average

Suppose

$$x_t = (1 + \beta L)\varepsilon_t \quad , \quad \varepsilon_t \text{ white} \quad |\beta|<1 \quad .$$

Then we have

$$P_{t-1}x_t = [L^{-1}(1 + \beta L)]_+ \left(\frac{1}{1+\beta L}\right)x_{t-1}$$

$$P_{t-1}x_t = \frac{\beta}{1+\beta L}\, x_{t-1}$$

We also have that for $k\geq 2$

$$P_{t-k}x_t = [L^{-k}(1 + \beta L)]_+ \left(\frac{1}{1+\beta L}\right) x_{t-1}$$

$$= 0 \quad ,$$

which can also be seen directly by projecting on $\{x_{t-k}, x_{t-k-1}, \ldots\}$ both sides of

$$x_t = (1 + \beta L)\varepsilon_t \quad .$$

First order moving average, autoregressive

Suppose we have

$$x_t = \left(\frac{1+aL}{1-\beta L}\right)\varepsilon_t \ , \quad \varepsilon_t \text{ white}, \ |a|<1, \ |\beta|<1 \quad .$$

We then have

$$P_{t-1}x_t = \left(\frac{L^{-1}(1+aL)}{(1-\beta L)}\right)_+ \left(\frac{1-\beta L}{1+aL}\right)x_{t-1} = \left(\frac{L^{-1}}{1-\beta L} + \frac{a}{1-\beta L}\right)_+ \left(\frac{1-\beta L}{1+aL}\right)x_{t-1}$$

$$P_{t-1}x_t = \left(\frac{\beta+a}{1-\beta L}\right)\left(\frac{1-\beta L}{1+aL}\right)x_{t-1}$$

$$P_{t-1}x_t = \{\frac{a+\beta}{1+aL}\}x_{t-1}$$

which expresses the forecast of x_t as a geometric distributed lag of past x's. The first order mixed moving average, autoregressive model for x_t thus provides a rationalization for the familiar "adaptive expectations" model. As we let $\beta \to 1$ (from below, in order to assure that the roots condition $|\beta|<1$ is met), $P_{t-1}x_t$ approaches

$$P_{t-1}x_t = \{\frac{1+a}{1+aL}\} x_{t-1}$$

which with $a<0$ is equivalent with Cagan's adaptive expectations scheme

$$P_{t-1}x_t = \{\frac{1-\lambda}{1-\lambda L}\} x_{t-1}$$

with $a=-\lambda$. Notice that as $\beta \to 1$ (from below), we approach the situation in which

$$(1-L)x_t = (1+aL),$$

so that the first difference of x_t follows a first order moving average. The parameter a must be negative in order that $\lambda>0$.

For the general case in which $k \geq 1$, we have

$$\begin{aligned} P_{t-k}x_t &= \left(\frac{L^{-k}(1+aL)}{1-\beta L}\right)_+ \left(\frac{1-\beta L}{1+aL}\right)x_{t-k} \\ &= \left(\frac{L^{-k}}{1-\beta L} + \frac{aL^{-k+1}}{1-\beta L}\right)_+ \left(\frac{1-\beta L}{1+aL}\right)x_{t-k} \\ &= \left(\frac{\beta^k}{1-\beta L} + \frac{a\beta^{k-1}}{1-\beta L}\right)\left(\frac{1-\beta L}{1+aL}\right)x_{t-k} = \frac{\beta^{k-1}(\beta+a)}{(1+aL)} x_{t-k} \end{aligned}$$

We can write this alternatively as

$$P_t x_{t+k} = \frac{\beta^{k-1}(\beta+a)}{1+aL} x_t \quad .$$

Notice that as $\beta \to 1$ (from below) we approach the situation in which

$$P_t x_{t+k} = \left(\frac{\beta+a}{1+aL}\right) x_t \quad ,$$

so that the same forecast is made for all horizons $k \geq 1$. In this sense, there is a well-defined concept of "permanent x." This was first pointed out in the economics literature by John F. Muth,* who showed that the hypothesis of rational expectations in conjunction with the model for income $(1-L)x_t = (1+aL)\varepsilon_t$ provides a rationalization both for the concept of permanent income and the geometric distributed lag formula that Milton Friedman had earlier used to estimate permanent income in empirical work.

*

Deriving a Moving Average Representation

The univariate prediction formulas given above assume that one has in hand a moving average representation for the covariance stationary, zero mean process $\{x_t\}$. Often, all that one has is the covariogram $c(\tau)$ of x from which the appropriate moving average representation must be calculated. To illustrate one method of finding the moving average coefficients, suppose that $c(\tau)$ is simply zero for $|\tau|>1$, so that only $c(0)$ and $c(1)$ are nonzero. It is apparent that x_t then has a first-order moving average representation

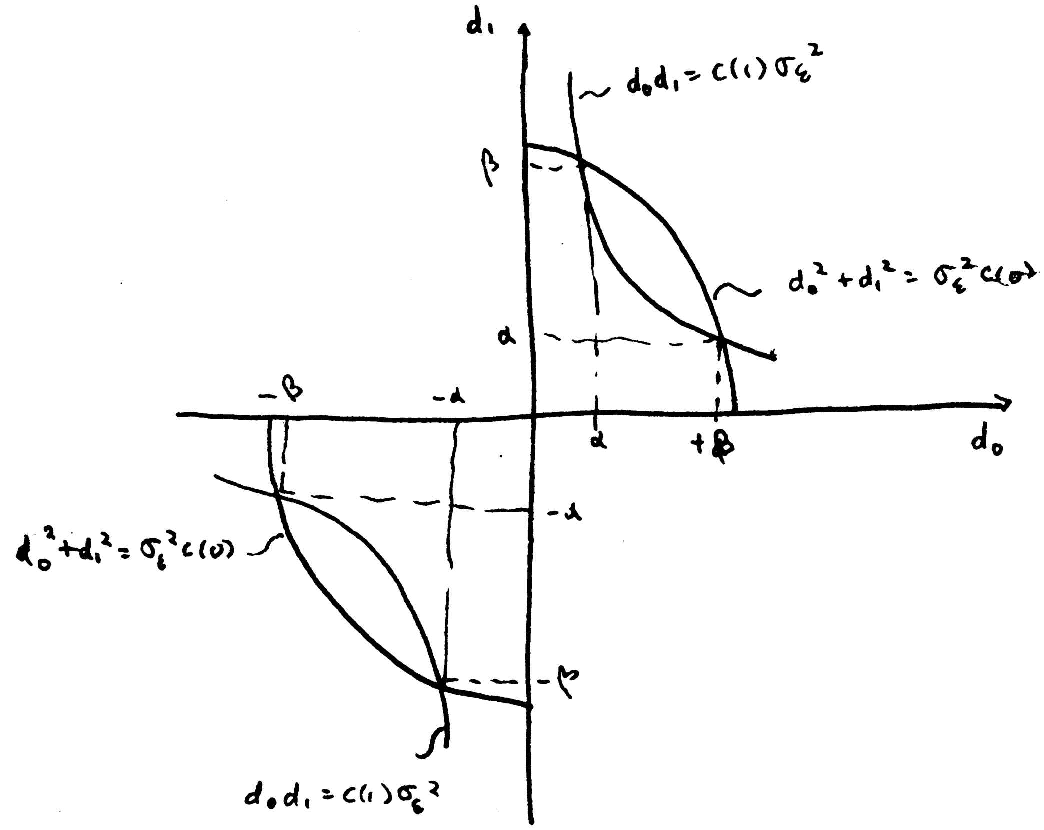
$$(60) \qquad x_t = d_0\varepsilon_t + d_1\varepsilon_{t-1}$$

where d_0 and d_1 are to be determined, and ε_t is required to be a white noise process of errors in predicting x_t from its own past. As we shall see, this latter condition must be imposed in order to determine the d's. For a process obeying (60) with $\{\varepsilon_t\}$ being a white noise with variance σ_ε^2, it is straightforward to calculate

$$(61) \qquad c(0) = (d_0^2 + d_1^2)\sigma_\varepsilon^2$$

$$c(1) = (d_0 d_1)\sigma_\varepsilon^2$$

Given the known values of $c(0)$ and $c(1)$ that characterize the x process, these are two (nonlinear) equations that can be solved for d_0 and d_1, given an assumed value for g_ε^2. The equations are graphed for fixed σ_ε^2 and $c(1)>0$ in Figure . In general, the two equations determine two pairs of solutions, one pair consisting of $d_0 = \alpha > \beta = d_1$ and $d_1 = \alpha > \beta = d_0$,



d_1
$d_0 d_1 = c(1)\sigma_\varepsilon^2$
β
$d_0^2 + d_1^2 = \sigma_\varepsilon^2 c(0)$
α
$-\beta$
$-\alpha$
α
$+\beta$
d_0
$-\alpha$
$d_0^2 + d_1^2 = \sigma_\varepsilon^2 c(0)$
$-\beta$
$d_0 d_1 = c(1)\sigma_\varepsilon^2$

where α and β are the positive scalars depicted in Figure ; the second pair is the reflection of the first pair in the negative quadrant. As σ_ε^2 varies, the solutions for d_0 and d_1 vary in a way easily determined from the graphs. We can forget about the solutions in the negative quadrant, since our discussion of Wold's theorem indicates that we want to choose $d_0=1$. Which of the two solutions with $d_0>0$ should be chosen? The answer comes from the condition that the derived ε_t process has to have a convergent series representation in terms of current and lagged x's. Suppose, for example, that we choose the solution for which $d_1>d_0$. We have

$$x_t = d_0\varepsilon_t + d_1\varepsilon_{t-1}$$

or

$$\text{(62)} \qquad \varepsilon_t = \left(\frac{1}{d_0}\right)x_t - \left(\frac{d_1}{d_0}\right)\varepsilon_{t-1}$$

so that ε_t cannot be expressed as a convergent series of lagged x's. That is, the backward solution of the above equation

$$\varepsilon_t = \frac{1}{d_0}\sum_{j=0}^{\infty}\left(\frac{-d_1}{d_0}\right)^j x_{t-j}$$

is not convergent because $\left|\left(\frac{-d_1}{d_0}\right)\right| > 1$. The forward solution of the difference equation (62) is "stable" if $d_1>d_0$. That is, as we saw earlier we can write

$$\varepsilon_t = \frac{1}{d_0+d_1L}\,x_t = \left(\frac{\left(\frac{1}{d_1}\right)L^{-1}}{1+\left(\frac{d_0}{d_1}\right)L^{-1}}\right)x_t$$

so that

$$\varepsilon_t = \frac{1}{d_1} \sum_{j=1}^{\infty} \left(\frac{d_0}{d_1}\right)^{j-1} x_{t+j}$$

which if $|d_0|<|d_1|$ expresses ε_t as a convergent (square summable) series of <u>future</u> x's. Thus, if $d_1>d_0$, the associated ε_t does not lie in the space spanned by current and lagged x's. However, if $d_0>d_1$, the associated ε_t process does lie in the space spanned by current and lagged x's,* which is the condition that will always result in choosing the correct roots of (61). The general principle is this: in selecting among the sequences $\{d_0,d_1,d_2,...\}$ that solve the equations that are the general counterparts of (61), choose the representation in which $d_0\sigma_\varepsilon^2$ is <u>maximal</u>. This selection is the unique one that makes $d_0\varepsilon_t$ the one-step ahead error in predicting x_t linearly from its own past; the ε_t's with this property are said to be the <u>fundamental</u> white noise process for x_t. Ordinarily, we normalize by choosing σ_ε^2 so that $d_0=1$. In this case ε_t equals the one-step ahead prediction error for x_t.

As a practical matter, solving the equations of the form (61) can be very tedious because they are highly nonlinear. A method of achieving an approximation to the moving average representation is to use $c(\tau)$ to calculate an autoregressive representation of some order n, i.e., to use the $c(\tau)$'s to fill out the elements of the least squares normal equations required to compute the a_i^n's in

---

* By an appropriate limiting argument it can be shown that ε_t lies in that space even if $d_0=d_1$.

$$x_t = \sum_{i=1}^{n} a_i^n x_{t-i}$$

where $E\varepsilon_t^n x_{t-i} = 0$ for $i=1,\ldots,n$. Then an approximation to the moving average lag operator $d(L)$ can be taken as

$$d_n(L) = (1 - \sum_{i=1}^{n} a_i^n L^i)^{-1} \quad .$$

By making n large enough, an arbitrarily good approximation* to $d_n(L)$ can be obtained.

---

* Arbitrarily good in the sense that the variance of the $\{\varepsilon_t^n\}$ process can be made as close as desired to the variance of $\{\varepsilon_t\}$ by making n large enough.

The Chain Rule of Forecasting

The law of iterated projections implies a recursion relationship that is sometimes very useful in a forecasting context. The relationship is known as Wold's "chain rule of forecasting." It shows how projections $P_t x_{t+k}$ for all $k \geq 2$ can be calculated from knowledge of the form of $P_t x_{t+1}$ alone.

Suppose that $\{x_t\}$ is a linearly indeterministic covariance stationary stochastic process for which

$$P_t x_{t+1} = \sum_{j=0}^{\infty} h_j x_{t-j} \quad , \quad \sum_{j=0}^{\infty} h_j^2 < \infty \quad .$$

It follows that

$$P_{t+k} x_{t+k+1} = h_0 x_{t+k} + h_1 x_{t+k-1} + \ldots + h_k x_t + h_{k+1} x_{t-1} + \ldots \quad .$$

Projecting both sides of this equation on $(x_t, x_{t-1}, \ldots)$ gives, via the law of iterated projections,

(63) $$P_t x_{t+k+1} = h_0 P_t x_{t+k} + h_1 P_t x_{t+k-1} + \ldots + h_{k-1} P_t x_{t+1} + \sum_{i=0}^{\infty} h_{k+i} x_{t-i} \quad .$$

This recursion relationship is the "chain rule of forecasting" which shows how to build up projections of x_t arbitrarily far into the future from knowledge alone of the formula for the one-step ahead projection.

To take an example, suppose that $\{x_t\}$ is a first order Markov process so that

$$P_t x_{t+1} = \lambda x_t \qquad |\lambda| < 1 \quad .$$

From application of (63) it follows that

$$P_t x_{t+j} = \lambda^j x_t \qquad j \geq 1 \quad .$$

Some Applications to Rational Expectations Models

Let us return to the example of Cagan's portfolio balance schedule, only now where we assume that m_t is a covariance stationary stochastic process and the price level now expected for next period is the linear least squares projection of P_{t+1} on information available at time t. We then have the difference equation

$$(64) \qquad m_t - p_t = \alpha P_t p_{t+1} - \alpha p_t \qquad \alpha<0$$

where $P_t p_{t+1}$ is the linear least squares forecast of p_{t+1} given information available at time t. Projecting the above equation on information available at time (t-1) gives

$$P_{t-1} m_t = \alpha P_{t-1} p_{t+1} + (1 - \alpha) P_{t-1} p_t$$

or

$$\left(B^{-1} + \frac{1-\alpha}{\alpha}\right) P_{t-1} p_t = \frac{1}{\alpha} P_{t-1} m_t$$

where $BP_{t-1} x_{t+j} \equiv P_{t-1} x_{t+j-1}$ and $B^{-1} P_{t-1} x_{t+j} \equiv P_{t-1} x_{t+j+1}$. Operating on both sides of the above equation by B gives

$$(1 - \frac{\alpha-1}{\alpha} B) P_{t-1} p_t = \frac{1}{\alpha} P_{t-1} m_t \quad .$$

As before, since $\alpha<0$ and $\frac{\alpha-1}{\alpha} >1$, we should solve this equation in the forward direction. Proceeding exactly as with our earlier calculations, we obtain the solution

$$P_{t-1}p_t = \frac{1}{1-\alpha}\left(\sum_{j=0}^{\infty}\left(\frac{\alpha}{\alpha-1}\right)^j B^j\right)P_{t-1}m_t$$

$$P_{t-1}p_t = \frac{1}{1-\alpha}\sum_{j=0}^{\infty}\left(\frac{\alpha}{\alpha-1}\right)^j P_{t-1}m_{t+j}$$

which is identical with our earlier solution with $\{x_t\}$ being replaced by $P_{t-1}\{x_t\}$ everywhere.

Now suppose that m_t has the moving average representation

$$m_t = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j}$$

where $\sum d_j^2 < \infty$ and ε_t is fundamental for m. Then we have, applying (59),

$$P_{t-1}p_t = \frac{1}{1-\alpha}\sum_{j=0}^{\infty}\left(\frac{\alpha}{1-\alpha}\right)^j \left[\frac{d(L)}{L^{j+1}}\right]_+ \varepsilon_{t-1}$$

$$= \frac{1}{1-\alpha}\left[\frac{d(L)}{L} + \frac{d(L)}{L^2}\left(\frac{\alpha}{1-\alpha}\right) + \frac{d(L)}{L^3}\left(\frac{\alpha}{1-\alpha}\right)^2 + \cdots\right]_+ \varepsilon_{t-1}$$

$$P_{t-1}p_t = \frac{1}{1-\alpha}\left[\left(\sum_{j=0}^{\infty}\left(\frac{\alpha}{1-\alpha}\right)^j L^{-(j+1)}\right)d(L)\right]_+ \varepsilon_{t-1}$$

Then the solution for $P_{t-1}p_t$ in terms of current and lagged m_t's is (using $m_t = d(L)\varepsilon_t$)

$$P_{t-1}p_t = \frac{1}{1-\alpha}\left[\sum_{j=0}^{\infty}\left(\frac{\alpha}{1-\alpha}\right)^j L^{-(j+1)} d(L)\right]_+ \frac{1}{d(L)}\, m_{t-1} \quad .$$

Substituting the above expression for $P_{t-1}p_t$ into (64) gives

$$m_t = (1-\alpha)p_t + \frac{1}{1-\alpha}\left[\sum_{j+0}^{\infty}\left(\frac{\alpha}{1-\alpha}\right)^j L^{-(j+1)} d(L)\right]_+ \frac{1}{d(L)} m_t$$

which expresses the stochastic process for p_t as a function of the exogenous stochastic process for m_t.

The preceding solution process is a constructive one. A quicker method of solution is the following one. Let us again assume that m has the moving average representation $m_t = d(L)\varepsilon_t$. Guess at a solution for p_t of the form*

$$p_t = v(L)\varepsilon_t \quad , \quad \sum_{j=0}^{\infty} v_j L^j = v(L)$$

Then equation (64) can be written

$$d(L)\varepsilon_t = (1-\alpha)v(L)\varepsilon_t + \alpha\left(\frac{v(L)}{L}\right)_+ \varepsilon_t$$

which implies

$$d(L) = (1-\alpha)v(L) + \alpha\left(\frac{v(L)}{L}\right)_+$$

an equation that can be used to solve for v(L) as a function of d(L) and α. Once v(L) has been determined, the solution for p_t can be written

$$p_t = v(L)\varepsilon_t = v(L)\frac{1}{d(L)} m_t \quad .$$

This method of solution was used by John F. Muth.**

---

* It will turn out that the ε_t's are fundamental for p, i.e., they are the one-step ahead prediction errors. This rationalizes the prediction formulas to be used.

** "Rational Expectations and the Theory of Price Movement," Econometrica, 1961.

Let us now consider the supply-demand example of section where x_t is now a covariance stationary, indeterministic random process with mean zero and moving average representation

$$x_t = d(L)\varepsilon_t$$

Our system is naturally modified to become

$$C_t = -\beta p_t \qquad \beta>0$$

$$Y_t = \gamma P_{t-1}p_t + x_t \qquad \gamma>0$$

$$I_t = \alpha(P_t p_{t+1} - p_t) \qquad \alpha>0$$

$$Y_t = C_t + I_t - I_{t-1}$$

where Y_t is production, C_t demand for consumption, and I_t holdings of inventories. Substituting the first three equations into the third gives

$$(\gamma+\alpha)P_{t-1}p_t + (\alpha+\beta)p_t = \alpha P_t p_{t+1} + \alpha p_{t-1} - x_t \quad . \tag{65}$$

Taking projections of both sides against information available at time (t-1) gives

$$\alpha P_{t-1}p_{t+1} - (\gamma+\beta+2\alpha)P_{t-1}p_t + \alpha P_{t-1}p_{t-1} = P_{t-1}x_t$$

or

$$(B^{-1} - \phi + B)P_{t-1}p_t = \frac{1}{\alpha} P_{t-1}x_t$$

where $B^{-1}P_{t-1}z_t \equiv P_{t-1}z_{t+1}$, $BP_{t-1}z_t \equiv P_{t-1}z_{t-1}$, and where $\phi = \frac{\beta+\gamma}{\alpha} + 2>0$.

Multiplying by B gives

$$(1 - \phi B + B^2)P_{t-1}p_t = \frac{1}{\alpha} P_{t-1}x_{t-1}$$

(66) $$(1 - \lambda B)(1 - \frac{1}{\lambda} B)P_{t-1}p_t = \frac{1}{\alpha} P_{t-1}x_{t-1}$$

where $|\lambda|<1$ satisfies $\lambda + \frac{1}{\lambda} = \phi$. Notice that

$$\frac{1}{(1-\lambda B)(1-\frac{1}{\lambda} B)} = \frac{\lambda}{\lambda - \frac{1}{\lambda}} \frac{1}{1-\lambda B} - \frac{\frac{1}{\lambda}}{\lambda - \frac{1}{\lambda}} \frac{1}{1-\frac{1}{\lambda} B}$$

To insure covariance stationarity of the solution, we need to insist that all lag distributions are square summable. Therefore we substitute

$\frac{1}{1 - \frac{1}{\lambda} B} = \frac{-\lambda B^{-1}}{1-\lambda B^{-1}}$, which gives

$$\frac{1}{(1-\lambda B)(1- \frac{1}{\lambda} B)} = \frac{\lambda}{\lambda - \frac{1}{\lambda}} \frac{1}{1-\lambda B} + \frac{1}{\lambda - \frac{1}{\lambda}} \frac{B^{-1}}{1-\lambda B^{-1}}$$

Multiplying both sides by $(1-\lambda B)$ gives

$$\frac{1-\lambda B}{(1-\lambda B)(1- \frac{1}{\lambda} B)} = \frac{\lambda}{\lambda- \frac{1}{\lambda}} + \frac{(1-\lambda B)B^{-1}}{(\lambda- \frac{1}{\lambda})(1-\lambda B^{-1})}$$

Operating on both sides of (66) with the preceding operator gives

$$(1-\lambda B)P_{t-1}p_t = \frac{\lambda}{\lambda- \frac{1}{\lambda}} \frac{1}{\alpha} P_{t-1}x_{t-1} + \frac{(1-\lambda B)B^{-1}}{(\lambda- \frac{1}{\lambda})(1-\lambda B^{-1})} \frac{1}{\alpha} P_{t-1}x_t \quad .$$

Assume that the set conditioning $P_{t-1}x_t$ includes x_{t-1}. Then we have $(1-\lambda B)P_{t-1}x_t = P_{t-1}x_t - \lambda P_{t-1}x_{t-1} = P_{t-1}(x_t-\lambda x_{t-1}) = P_{t-1}(1-\lambda L)x_t$. Substituting this into the above equation gives

$$(1-\lambda B)P_{t-1}P_t = \frac{\lambda}{\lambda - \frac{1}{\lambda}} \frac{1}{\alpha} x_{t-1} + \frac{1}{\alpha} \frac{1}{\lambda - \frac{1}{\lambda}} \frac{1}{1-\lambda B^{-1}} P_{t-1}(1-\lambda L)x_t \quad ,$$

which is a solution for $P_{t-1}p_t$. This solution suggests that the solution for p_t is given by

$$p_t - \lambda p_{t-1} = \frac{1}{\alpha} \frac{\lambda}{\lambda - \frac{1}{\lambda}} x_{t-1} + \frac{1}{\alpha} \frac{\lambda}{\lambda - \frac{1}{\lambda}} \frac{1}{1-\lambda B^{-1}} P_t(1-\lambda L)x_t$$

which can be rearranged to read

$$(67) \qquad p_t - \lambda p_{t-1} = -\lambda \sum_{i=0}^{\infty} \lambda^i P_t(\frac{1}{\alpha} x_{t+i})$$

That (67) is a solution can be verified by direct substitution into (65). We can reduce (67) further by eliminating $P_t x_{t+i}$ via the Wiener-Kolmogorov formula to get

$$p_t - \lambda p_{t-1} = \left(\frac{-\lambda}{\alpha} \sum_{i=0}^{\infty} \lambda^i \left[\frac{d(L)}{L^i}\right]_+ \right) \frac{1}{d(L)} x_t$$

$$= \frac{-\lambda}{\alpha} \left[\frac{1}{1-\lambda L^{-1}} d(L)\right]_+ \frac{1}{d(L)} x_t \quad .$$

Fourier Analysis of Data

To motivate further the interpretation of the spectrum as a decomposition of variance by frequency, suppose that we have T observations on y_t, t=1, 2, ..., T. Suppose for convenience that Y is an even number (assuming that it is odd would require some minor modifications in some of the formulas that follow). We consider computing the following regression of y_t on sine and cosine functions of angular frequency $w_j=2\pi j/T$ where j=0, 1, ..., T/2:

(68) $$y_t = \sum_{k=0}^{T/2} \hat{\alpha}(w_k)\cos w_k t + \sum_{k=1}^{T/2-1} \hat{\beta}(w_k)\sin w_k t$$

where $w_k = 2\pi k/T$. There are T observations and T dependent variables in this regression, which means that the regression will fit perfectly provided that the regressors are linearly independent, as they are. Indeed, the regressors are pairwise orthogonal. Thus, recall that

$$\cos \lambda = (e^{i\lambda} + e^{-i\lambda})/2$$

$$\sin \lambda = (e^{i\lambda} - e^{-i\lambda})/2i.$$

Now use these equalities to write

(69) $$\sum_{t=1}^{T} \cos \frac{2\pi j}{T} t\left[\cos \frac{2\pi k}{T} t + i \sin \frac{2\pi k}{T} t\right]$$

$$= \sum_{t=1}^{T} \frac{1}{2}\left(e^{i\frac{2\pi j}{T}t} + e^{-i\frac{2\pi j}{T}t}\right)\left(e^{i\frac{2\pi k}{T}t}\right)$$

$$= \frac{1}{2}e^{i\frac{2\pi(j+k)}{T}} \sum_{t=0}^{T-1} e^{i\frac{2\pi(j+k)t}{T}} + \frac{1}{2} e^{i\frac{2\pi(k-j)}{T}} \sum_{t=0}^{T-1} e^{i\frac{2\pi(k-j)t}{T}}$$

$$= \begin{cases} \frac{1}{2}e^{i\frac{2\pi(j+k)}{T}} \left[\dfrac{1-e^{i2\pi(j+k)}}{1-e^{i\frac{2\pi(j+k)}{T}}}\right] + \frac{1}{2}e^{i\frac{2\pi(k-j)}{T}} \left[\dfrac{1-e^{i2\pi(k-j)}}{1-e^{i\frac{2\pi(k-j)}{T}}}\right] & 0 \leq k \neq j \leq \frac{1}{2} T \\ \frac{1}{2}e^{i\frac{2\pi(j+k)}{T}} \dfrac{1-e^{i2\pi(j+k)}}{1-e^{i\frac{2\pi(j+k)}{T}}} + \frac{1}{2}T, & 0 < k = j < \frac{1}{2} T \\ T & k = j = 0, \frac{1}{2} T \end{cases}$$

$$= \begin{cases} 0 \text{ (because } e^{i2\pi(j+k)} = 1) & 0 \leq k \neq j \leq \frac{1}{2}T \\ \frac{1}{2}T & 0 < k = j < \frac{1}{2}T \\ T & k = j = 0, \frac{1}{2}T \end{cases}$$

Equating the real and imaginary parts, respectively, of the first line in this equation with the last gives

$$\text{(70)} \qquad \sum_{t=1}^{T} \cos \frac{2\pi j}{T} t \cos \frac{2\pi k}{T} t = \begin{cases} 0 & 0 \leq k \neq j \leq \frac{1}{2}T \\ \frac{1}{2}T & 0 < k = j < \frac{1}{2}T \\ T & k = j = 0, \frac{1}{2}T \end{cases}$$

and

$$\text{(71)} \qquad \sum_{t=1}^{T} \cos \frac{2\pi j}{T} t \sin \frac{2\pi k}{T} t = 0 \qquad k, j = 0, 1, \ldots, \frac{1}{2}T.$$

A similar argument shows that

$$\text{(72)} \qquad \sum_{t=1}^{T} \sin \frac{2\pi j}{T} t \sin \frac{2\pi k}{T} t = \begin{cases} 0 & 0 \leq k \neq j \leq \frac{1}{2}T \\ \frac{1}{2}T & 0 < k = j < \frac{1}{2}T \\ 0 & k = j = 0, \frac{1}{2}T. \end{cases}$$

Taken together, equalities (70), (71), and (72) show that the regressors in (68) are mutually orthogonal. Notice that setting j=0 in (70) and (72) gives

$$\sum_{t=1}^{T} \cos \frac{2\pi k}{T} t = 0 = \sum_{t=1}^{T} \sin \frac{2\pi k}{T} t, \; k=1, 2, \ldots, T/2.$$

Where the regressors are mutually orthogonal, as they are in (68), the least squares estimator of the multiple regression coefficients is identical with the vector of simple least squares regression coefficients. These are given by

$$\hat{\alpha}(w_k) = \frac{\sum_{t=1}^{T} y_t \cos w_k t}{\sum_{t=1}^{T} \cos^2 w_k t} \qquad k = 0, 1, \ldots, T/2$$

$$\hat{\beta}(w_k) = \frac{\sum_{t=1}^{T} y_t \sin w_k t}{\sum_{t=1}^{T} \sin^2 w_k t} \qquad k = 1, 2, \ldots, T/2-1$$

Using (70), (71), and (72), the above can be simplified to

$$\hat{\alpha}(w_0) = \frac{\sum_{t=1}^{T} y_t}{T}$$

$$\hat{\alpha}(w_{T/2}) = \frac{1}{T}\sum_{t=1}^{T} y_t (-1)^t$$

$$\hat{\alpha}(w_k) = \frac{2}{T}\sum_{t=1}^{T} y_t \cos w_k t, \qquad k = 1, 2, \ldots, T/2-1$$

$$\hat{\beta}(w_k) = \frac{2}{T}\sum_{t=1}^{T} y_t \sin w_k t, \qquad k = 1, 2, \ldots, T/2-1.$$

Since (68) represents a regression of T observations on y against T orthogonal independent variables (which guarantees that the X'X matrix of the linear statistical model is of full rank), we know that the regression fits the data exactly, i.e., it gives a perfect fit. So what we have achieved is a decomposition of $y_t\{t=1, \ldots, T\}$ into a weighted sum of sine and cosine terms of angular frequencies $w_k = \frac{2\pi k}{T}$, $k=0, \ldots, T/2$. The least squares regression coefficients $\hat{\alpha}(w_k)$ and $\hat{\beta}(w_k)$ give a measure of how important the various frequencies are in composing the series y_t. To make this more precise, notice that from (68), the sample variance of the y's can be written

$$\frac{1}{T}\sum_{t=1}^{T}\left(y_t-\frac{\sum_{t=1}^{T}y_t}{T}\right)^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t-\hat{\alpha}(w_0))^2$$

$$= \frac{1}{T}\left\{\sum_{k=1}^{T/2-1}\hat{\alpha}(w_k)^2\sum_{t=1}^{T}\cos^2(w_k t) + \sum_{k=1}^{T/2-1}\hat{\beta}(w_k)^2\sum_{t=1}^{T}\sin^2(w_k t)\right.$$

$$\left.+\ \hat{\alpha}(w_{T/2})^2\sum_{t=1}^{T}\cos^2(w_{T/2}t)\right\},$$

which follows by virtue of the orthogonality of sines and cosines of different frequencies. From our earlier calculations of $\sum_{t=1}^{T}\cos^2 w_k t$ and $\sum_{t=1}^{T}\sin^2 w_k t$, the above equation becomes

(73) $$\frac{1}{T}\sum_{t=1}^{T}\left(y_t-\frac{\sum y_t}{T}\right)^2 = \frac{1}{T}\left\{T/2\sum_{k=1}^{T/2-1}[\hat{\alpha}^2(w_k)+\hat{\beta}^2(w_k)]\right.$$

$$\left.+\ T\hat{\alpha}^2(w_{T/2})\right\} = \frac{1}{2}\sum_{k=1}^{T/2-1}[\hat{\alpha}^2(w_k)+\hat{\beta}^2(w_k)] + \hat{\alpha}^2(w_{T/2}) .$$

Thus, the term $1/2[\hat{\alpha}^2(w_k)+\hat{\beta}^2(w_k)]$ measures the contribution of sine and cosine terms of frequency w_k to the sample variance of y. Equation (73) is an example of Parseval's relation.

An equivalent but more compact version of the preceding decomposition is provided by the exponential Fourier series representation:

(74) $$y_t = \sum_{j=-T/2+1}^{T/2}\gamma_j e^{-i\frac{2\pi j}{T}t} \qquad t=1,\ 1,\ \ldots,\ T$$

which provides an exact representation of y_t, t=1, ..., T. We assert that the γ_h's are given by

(75) $$\gamma_h = \frac{1}{T}\sum_{t=1}^{T}y_t e^{+i\frac{2\pi h}{T}t}$$

This can be verified by substituting (74) into the above equality to get

$$\gamma_h = \frac{1}{T}\sum_{j=-T/2+1}^{T/2} \gamma_j \sum_{t=1}^{T} e^{i\frac{2\pi(h-j)}{T}t} .$$

But

$$\sum_{t=1}^{T} e^{i\frac{2\pi(h-j)t}{T}} = \begin{cases} T & \text{for } j=h \\ e^{i\frac{2\pi(h-j)}{T}} \left[\frac{1-e^{i2\pi(h-j)}}{1-e^{i\frac{2\pi(h-j)}{T}}}\right] = 0 & \text{for } j\neq h \end{cases}$$

Thus, we verify that

$$\frac{1}{T}\sum_{t=1}^{T} y_t e^{i\frac{2\pi ht}{T}} = \gamma_h$$

The list of the $\gamma_h \equiv \gamma(w_h)$ by frequency is called the finite Fourier transform of y_t, t=1, ..., T. To match this up with our earlier work write

$$\gamma_h = \frac{1}{T}\sum_{t=1}^{T} y_t \cos\frac{2\pi ht}{T} + i\,\frac{1}{T}\sum_{t=1}^{T} y_t \sin\frac{2\pi ht}{T}$$

$$= \alpha_h + i\beta_h$$

where

$$\alpha_h = \frac{1}{T}\sum_{t=1}^{T} y_t \cos\frac{2\pi ht}{T}$$

and

$$\beta_h = \frac{1}{T}\sum_{t=1}^{T} y_t \sin\frac{2\pi ht}{T} .$$

Substituting

$$\gamma_h = \alpha_h + i\beta_h$$

into (74) and writing

$$\cos \frac{2\pi j}{T} t - i \sin \frac{2\pi j}{T} t$$

for

$$e^{-i\frac{2\pi j}{T} t}$$

and noting that $\gamma_h = \overline{\gamma}_{-h}$ so that $\alpha_h = \alpha_{-h}$, $\beta_h = -\beta_h$, gives

$$y_t = \gamma_0 + \sum_{j=1}^{T/2} \alpha_j (\cos \frac{2\pi jt}{T} - i \sin \frac{2\pi j}{T} t)$$

$$+ \sum_{j=1}^{T/2-1} \alpha_j (\cos \frac{-2\pi jt}{T} - i \sin \frac{-2\pi j}{T} t)$$

$$+ i \sum_{j=1}^{T/2} \beta_j (\cos \frac{2\pi jt}{T} - i \sin (\frac{2\pi j}{T} t)$$

$$- i \sum_{j=1}^{T/2-1} \beta_j (\cos (\frac{-2\pi jt}{T}) + i \sin (\frac{-2\pi jt}{T}))$$

(76) $$y_t = \alpha_0 + 2 \sum_{j=1}^{T/2-1} \alpha_j \cos \frac{2\pi jt}{T} + \alpha_{T/2} \cos \frac{2\pi j T/2}{T}$$

$$+ 2 \sum_{j=1}^{T/2-1} \beta_j \frac{\sin 2\pi j}{T} t.$$

Comparing α_j and β_j in (76) with our earlier least squares estimates, we have

$$\alpha_0 = \hat{\alpha}(w_0)$$

$$\alpha_k = \frac{1}{2}\hat{\alpha}(w_k), \qquad k = 1, 2, \ldots, T/2-1$$

$$\alpha_{T/2} = \hat{\alpha}(w_{T/2})$$

$$\beta_k = \frac{1}{2}\hat{\beta}(w_k)$$

Thus, the real and imaginary parts of $\gamma_h = \alpha_h + i\beta_h$ are (apart from a scalar for k=1, ..., T/2-1) the regression coefficients in (68).

A "natural" measure of the importance of the cosine and sine waves of frequency w_k in composing y_t is the squared amplitude of $\gamma_k \equiv \gamma(w_k)$

$$\gamma(w_k)\overline{\gamma(w_k)} = |\gamma(w_k)|^2$$

$$= (\alpha(w_k)+i\beta(w_k))(\alpha(w_k)-i\beta(w_k))$$

$$= \alpha^2(w_k) + \beta^2(w_k).$$

The higher is this quantity, the larger are the weights placed on the sine and cosine of frequency w_k in (51) in making up y_t. The quantity $|\gamma(w_k)|^2$ is called the periodogram ordinate at frequency w_k, and turns out to provide a basis for estimating the spectrum at w_k. The relationship between the spectrum and the periodogram ordinates $\alpha^2+\beta^2$ provides one illuminating way of depicting the spectrum as a decomposition of variance by frequency.

To establish the relationship between the periodogram ordinates and the spectrum and cross spectrum, suppose that we have observations on two jointly covariance stationary stochastic processes y_t and x_t for $t = -T+1, -T+2, \ldots, -1, 0, 1, \ldots, T$. Assume that y_t and x_t have zero means. Then we compute the Fourier transforms

$$x(w_k) = \frac{1}{2T} \sum_{t=-T+1}^{T} x_t e^{-iw_k t}$$

$$y(w_k) = \frac{1}{2T} \sum_{t=-T+1}^{T} y_t e^{-iw_k t} \quad , \qquad w_k = \frac{\pi k}{T} \text{ , } k=0,\ 1,\ \ldots,\ T\ .$$

Consider now the <u>cross-periodogram</u> ordinates defined by

$$2T\ y(w_k)x^*(w_n) = \frac{1}{2T} \sum_{t=-T+1}^{T} \sum_{s=-T+1}^{T} y_t x_s e^{-iw_k t}\ e^{iw_n s}$$

Letting $s=t-\tau$ so that $\tau=t-s$, we have

$$2T\ y(w_k)x^*(w_n) = \frac{1}{2T} \sum_{t=-T+1}^{T} \sum_{\tau=t-T}^{t+T-1} y_t x_{t-\tau} e^{-iw_k t}\ e^{iw_n(t-\tau)}$$

$$= \frac{1}{2T} \sum_{t=-T+1}^{T} \sum_{\tau=t-T}^{t+T-1} y_t x_{t-\tau} e^{-iw_n\tau}\ e^{-i(w_k-w_n)t}$$

Taking expected values, we have

(77) $$E2Ty(w_k)x^*(w_n) = \frac{1}{2T} \sum_{t=-T+1}^{T} e^{-i(w_k-w_n)t} \sum_{\tau=t-T}^{t+T-1} e^{-iw\tau} C_{yx}(\tau)$$

For $b>a$, b and a integers, we have

$$\sum_{j=a}^{k} e^{i\lambda j} = \begin{cases} e^{i\lambda(b+a)/2}\ \dfrac{\sin\ \lambda(b-a+1)/2}{\sin\ (\lambda/2)} & \lambda\neq 0 \\ b - a + 1 & \lambda=0 \end{cases}$$

This is obvious for $\lambda=0$. For $\lambda\neq 0$, we have

$$\sum_{j=a}^{b} e^{i\lambda j} = \frac{e^{i\lambda a} - e^{i\lambda(b+1)}}{1 - e^{i\lambda}}$$

$$= \frac{e^{i\lambda(a+b+1)/2}}{e^{i\frac{\lambda}{2}}} \frac{(e^{-i\lambda(b-a+1)/2} - e^{i\lambda(b-a+1)/2})}{e^{-i\frac{\lambda}{2}} - e^{+i\frac{\lambda}{2}}}$$

$$= e^{i\lambda(a+b)/2}\left[\frac{\sin\ (\lambda(b-a+1)/2)}{\sin\ (\frac{\lambda}{2})}\right]$$

Therefore, for $w_k = w_n$, we have

$$(78) \qquad \sum_{t=-T+1}^{T} e^{i(w_n - w_k)t} = 2T \quad .$$

For $w_k \neq w_n$, we have

$$(79) \qquad \sum_{t=-T+1}^{T} e^{i(w_n - w_k)t} = e^{i(w_n - w_k)/2} \frac{\sin\ (w_n - w_k)T}{\sin\left(\frac{w_n - w_k}{2}\right)} \quad ,$$

which is bounded in absolute value by $1/\sin\left(\frac{w_n - w_k}{2}\right)$ for all T. Substituting (78) and (79) into (77) and taking the limit as T goes to infinity, we have

$$(80) \qquad E2Ty(w_k)x^*(w_n) = \begin{cases} \sum_{\tau=-\infty}^{\infty} e^{-iw_k\tau} C_{yx}(\tau) & w_k = w_n \\ 0 & w_k \neq w_n \end{cases}$$

or

$$E2Ty(w_k)x^*(w_k) = \begin{cases} g_{yx}(e^{-iw_k}) & w_k = w_n \\ 0 & w_k \neq w_n \end{cases}$$

In other words, for large enough T, $E2Ty(w_k)x^*(w_k)$ approaches closely to the cross spectrum, while the ordinates $y(w_k)$ and $x^*(w_h)$ are asymptotically orthogonal if $w_k \neq w_n$.

For the special case in which $x_t \equiv y_t$, the above results show that as $T \to \infty$,

$$E2Tx(w_k)x^*(w_n) \to \begin{cases} g_x(e^{-iw_k}) & w_k = w_n \\ 0 & w_k \neq w_n \end{cases}$$

This shows that the periodogram ordinates $2T|x(w_k)|^2$ are asymptotically unbiased estimators of the spectrum at frequency w_k. Let us denote the periodogram ordinate by

$$I_T(w_k) \equiv 2T|x(w_k)|^2 \quad .$$

By using (80) and performing a few additional calculations, the following properties of the periodogram ordinates could be established.* Assume that $\{x_t\}$ obeys the normal probability law. Then we have that for k not equal to zero or T,

$$\frac{2I_T(w_k)}{g_x(e^{-iw_k})}$$

is distributed asymptoticaLly as chi square with <u>two</u> degrees of freedom. For k equal to zero or T,

$$\frac{I_T(w_k)}{g_x(e^{-iw_k})}$$

---

* See Koopmans, pp. 260-265.

is distributed asymptotically as chi square with one degree of freedom. Since a chi-square variate with r degrees of freedom has mean r and variance 2r, it follows that (asymptotically)

$$EI_T(w_k) = g_x(e^{-iw_k})$$

$$\text{(81)} \qquad \text{var } I_T(w_k) = (g_x(e^{-iw_k}))^2 \qquad k \neq 0, T$$

$$\text{var } I_T(w_k) = 2(g_x(e^{-iw_k}))^2 \qquad k = 0, T \quad .$$

Further, it can be shown as an implication of (80) that periodogram ordinates are asymptotically independent, so that $I(w_k)$ is asymptotically independent of $I(w_h)$ for $w_k \neq w_h$.

From (81) we see that the (asymptotic) variance of the periodogram ordinates does not depend on T, and in particular does not decrease with increases in sample size T. Therefore, though $I_T(w_k)$ is an (asymptotically) unbiased estimator of $g_x(e^{-iw})$, it is not consistent; i.e. there is no tendency for the variance of $I_T(w_k)$ around $g_x(e^{-iw_k})$ to decrease as $T \to \infty$. This is the reason that raw periodogram ordinates $I(w_k)$ are regarded as noisy estimates of the spectrum.

In applied work, the spectrum and cross-spectrum are estimated by first calculating the periodogram and cross-periodogram ordinates. Then the assumption is adopted that the population spectrum and cross-spectrum are "smooth" functions of w. (To make these assumptions approximately correct, the data are typically filtered to give series with approximately locally flat spectra and cross-spectra.) Then the spectrum and cross-spectrum are estimated by taking some sort of moving average of periodogram ordinates across frequencies. Since the periodogram ordinates are asymptotically orthogonal, this averaging reduces the sampling variability of the resulting estimates. In effect, different spectral estimators differ only in the form of the moving average they apply.*

The Cramér Representation**

We have seen that the spectrum of x_t represents an orthogonal decomposition by frequency of the variance of x_t. Suppose we select a set of points $0=w_1<w_2<\ldots<w_{n+1}=\pi$. We then form $B_i(L)$ to satisfy

$$B_i(e^{-iw}) = \begin{cases} 1 & w\varepsilon[-w_{i+1}-w_i]\cap[w_i,w_{i+1}] \\ 0 & w\notin[-w_{i+1},-w_i]\cap[w_i,\ w_{i+1}] \end{cases}$$

Define $x_{it} = B_i(L)x_t$. Then we have that

$$x_t = \sum_{i=1}^{n} x_{it}$$

where $Ex_{it}\cdot x_{js}=0$ for $i\neq j$ and all s and t. So for any finite n, we are able to decompose x_t by frequency into n orthogonal processes, where the spectrum of x_{it} equals the spectrum of x_t wherever the spectrum of x_{it} is not zero.

This section in effect addresses the question: what happens

---

* See Koopmans and T. W. Anderson.

** This section is optional. It follows the treatment in Papoulis [pp. 468-472].

when we drive n to infinity in the above construction? It is perhaps natural to conjecture that x_t can be represented in a form

$$x_t = \int_{-\pi}^{\pi} e^{+iwt} dF(w)$$

where F(w) is a stochastic process with certain strong orthogonality properties inherited from those of the x_{it}'s. This conjecture is correct and is the intuition underlying the Cramér representation possessed by all covariance stationary stochastic processes.

Let $\{x_t\}_{t=-\infty}^{\infty}$ be a covariance stationary stochastic process with mean zero. It would be tempting to try to compute the Fourier transform of the $\{x_t\}$ process according to

$$\text{(82)} \qquad f(w) = \sum_{t=-\infty}^{\infty} x_t e^{-iwt} \qquad w\varepsilon[-\pi,\pi]$$

and thereby obtain a new stochastic process f(w); that is, essentially (82) would be used to define f(w) for each realization of $\{x_t\}_{t=-\infty}^{\infty}$, thereby creating a probability distribution on f(w)'s, which are thus random functions defined on $[-\pi,\pi]$. Unfortunately, however, the right side of (82) is <u>not</u> in general well defined. Fundamentally, this is because realizations of a covariance stationary stochastic process $\{x_t\}$ need <u>not</u> in general be square summable (i.e. satisfy $\sum_{t=-\infty}^{\infty} |x_t|^2 < \infty$), so that there is no guarantee that the infinite sum on the right side of (82) is well defined.

However, the "generalized Fourier transform" of the stochastic process x_t <u>is</u> well defined. In particular, let us define a stochastic process G(w), $w\varepsilon[-\pi,\pi]$ as follows. We set $G(-\pi)=0$ and for $\pi \geq w_1 > w_2 \geq -\pi$

we define

$$(83) \qquad G(w_1) - G(w_2) = \sum_{t=-\infty}^{\infty} \frac{e^{-iw_1 t} - e^{-iw_2 t}}{-it} x_t$$

The complex-valued random process G(w) thus defined is called the "generalized Fourier transform" of the stochastic process x_t; G(w) is a stochastic process or "random function" defined on w in $[-\pi,\pi]$, because it is a function of the stochastic process $\{x_t\}$. The distribution of G(w) is traced out as the sequence $\{x_t\}_{t=-\infty}^{\infty}$ varies from realization to realization. The generalized Fourier transform of the x process is well defined even where the right side of (82) is not well defined. To indicate why, set $w_1=w+\varepsilon$ and $w_2=w-\varepsilon$ to obtain

$$(83') \qquad G(w+\varepsilon) - G(w-\varepsilon) = \sum_{t=-\infty}^{\infty} e^{-iwt}\left(\frac{e^{it\varepsilon} - e^{-it\varepsilon}}{it}\right) x_t$$

$$= \sum_{t=-\infty}^{\infty} e^{-iwt} \frac{2 \sin \varepsilon t}{t} x_t$$

Thus $G(w+\varepsilon) - G(w-\varepsilon)$ is the ordinary Fourier transform of x_t multiplied by $\frac{2 \sin \varepsilon t}{t}$. The function $2 \sin \varepsilon t/t$ goes to zero rapidly as $|t|\to\infty$. Heuristically, this permits $\frac{2 \sin \varepsilon t}{t}\cdot x_t$ to satisfy the square summability condition necessary for the Fourier transform to be well defined even where $\{x_t\}$ fails to be square summable. This is what underlies the fact (which we won't prove) that the generalized Fourier transform G(w) is well defined, i.e. the infinite sum on the right side of (83) converges in mean square.

Now divide both sides of (83') by $2\varepsilon, \varepsilon>0$, to obtain

(84) $$\frac{G(w+\varepsilon) - G(w-\varepsilon)}{2\varepsilon} = \sum_{t=-\infty}^{\infty} e^{-iwt}\left(\frac{\sin \varepsilon t}{\varepsilon t}\right)x_t \quad .$$

So $\frac{G(w+\varepsilon) - G(w-\varepsilon)}{2\varepsilon}$ is the ordinary Fourier transform of $\frac{\sin \varepsilon t}{\varepsilon t} x_t$. The function $\frac{\sin \varepsilon t}{\varepsilon t}$ is plotted in Figure 7. From l'Hospital's rule we have that

$$\lim_{t\to 0} \frac{\sin \varepsilon t}{\varepsilon t} = \lim_{t\to 0} \frac{\varepsilon \cos \varepsilon t}{\varepsilon} = 1 \quad .$$

The fact that $\frac{\sin \varepsilon t}{\varepsilon t} \to 0$ as $|t| \to \infty$ is what makes the infinite sum on the right side of (84) well defined.

We now formally apply the inversion formula to (84) to obtain

$$\frac{\sin \varepsilon t}{\varepsilon t} x_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{iwt} \frac{G(w+\varepsilon) - G(w-\varepsilon)}{2\varepsilon} dw$$

Letting $\varepsilon \to 0$, the left side approaches x_t, which we write as

$$x_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{+iwt} dG(w) \quad ,$$

which is the Cramér representation for the process x_t.
The "integral" on the right is defined as follows as a "mean square limit." Let P_n be a "partition" of the interval $[-\pi,\pi]$, i.e. P_n is a collection of points $w_i, i=1,\ldots,n$

$$P_n = [w_1, w_2, \ldots, w_n]$$

where $-\pi = w_1 < w_2 < \ldots < w_n =$. Let the "norm" of the partition be Δ;

$$\Delta_n = \max_{i=2,\ldots,n} (w_i - w_{i-1}) \quad .$$

Let $\{P_n\}_{n=2}^{\infty}$ be a sequence of partitions with $\Delta^n \to 0$ as $n \to \infty$. For each n let θ_i be points satisfying $w_{i-1} < \theta_i < w_i$. Then it can be shown that

$$(86) \qquad \lim_{n\to\infty} E\left|x_t - \frac{1}{2\pi}\sum_{i=1}^{n} e^{i\theta_i t}\,(G(w_i) - G(w_{i-1}))\right|^2 = 0 ,$$

i.e. the sequence of approximating sums $\frac{1}{2\pi}\sum_{i=1}^{n} e^{i\theta_i t}(G(w_i) - G(w_{i-1}))$ converges to x_t in the mean square sense that the variance of the difference between x_t and the approximating sum approaches zero. The notation

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{+iwt}\,dG(w)$$

is intended to denote the mean square limit of the sequence of approximating sums in (86).

Equation (85) is the "spectral representation" or Cramér representation for the stochastic process $\{x_t\}$. The random function G(w) is intimately related to the spectrum, as we shall now show.

To interpret G(w), let us return to a version of the "band pass" system we considered above. In particular, let

$$(87) \qquad h_{w_1 w_2}(e^{-iw}) = \begin{cases} 2\pi & w_2 < w < w_1 \\ 0 & w \notin [w_2, w_1] \end{cases}$$

The distributed lag weights $h_{w_1 w_2}(j)$ corresponding to (87) are, from the inversion formula, given by

$$(88) \qquad h_{w_1 w_2}(j) = \frac{2\pi}{2\pi}\int_{w_2}^{w_1} e^{iwj}\,dw = \frac{e^{iw_1 j} - e^{iw_2 j}}{ij} \quad .$$

With $\{x_t\}$ as the input to the system with these distributed lag weights, the output $y_{w_1w_2}(t)$ is given by

$$y_{w_1w_2}(t) = \sum_{j=-\infty}^{\infty} \frac{e^{iw_1j} - e^{iw_2j}}{ij} x_{t-j}$$

Adopting the change of variable $\tau=t-j$ so that $j=t-\tau$, we have

$$y_{w_1w_2}(t) = \sum_{\tau=-\infty}^{\infty} \frac{e^{iw_1(t-\tau)} - e^{iw_2(t-\tau)}}{i(t-\tau)} x_\tau \quad .$$

For $t=0$ we therefore have

$$(89) \qquad y_{w_1w_2}(0) = \sum_{\tau=-\infty}^{\infty} \frac{e^{-iw_1\tau} - e^{-iw_2\tau}}{-i\tau} x_\tau \quad ,$$

$$\equiv G(w_1) - G(w_2)$$

where the last equality is a repetition of the definition (83). Thus, the increment $G(w_1) - G(w_2)$ in $G(w)$ has the following interpretation: it is the random variable $y_{w_1w_2}(0)$ that is derived by applying the "bandpass" filter $h_{w_1w_2}(j)$ to the x_t process and evaluating the resulting stochastic process at time $t=0$.

The interpretation (89) is useful in establishing the properties of the random process $G(w)$. The first property is:

If $w_1>w_2$, then

$$E\{|G(w_2) - G(w_1)|^2\} = 2\pi\int_{w_2}^{w_1} g_x(e^{-iw})dw \quad .$$

To prove this, consider applying the bandpass filter $h_{w_1w_2}(j)$ defined by

(88) to the x_t process. From (35) we then know that the spectrum of the output*

$$g_x(e^{-iw})|h_{w_1w_2}(e^{-iw})|^2 = \begin{cases}(2\pi)^2 g_x(e^{-iw}), & w_2<w<w_1 \\ 0 & w\notin[w_2,w_1]\end{cases} ,$$

where $g_x(e^{-iw})$ is the spectrum of x. Then the variance of $y_{w_1w_2}(t)$, which equals the variance of $y_{w_1w_2}(0)$ by covariance stationarity, equals

$$E\{|y_{w_1w_2}(0)|^2\} = \frac{2\pi^2}{2\pi}\int_{w_2}^{w_1} g_x(e^{-iw})dw \quad .$$

But from (89) we have

$$E\{|y_{w_1w_2}(0)|^2\} = E\{|G(w_1) - G(w_2)|^2\} \quad .$$

The second property of G(w) is that it is a process with "orthogonal increments." That is, let $\pi>w_1>w_2\geq w_3>w_4\geq-\pi$. Form the two bandpass systems with frequency response functions

$$h_{w_1w_2}(e^{-iw}) = \begin{cases}2\pi & w_2<w<w_1 \\ 0 & 0,\ w\notin[w_2,w_1]\end{cases}$$

$$h_{w_3w_4}(e^{-iw}) = \begin{cases}2\pi & w_4<w<w_3 \\ 0 & 0,\ w\notin[w_4,w_3]\end{cases} \quad .$$

With x_t as the input to the systems with filters described by (88), we obtain outputs $y_{w_1w_2}(t)$ and $z_{w_3w_4}(t)$. From (22), it follows that their

---

* Notice that the output is a complex-valued stochastic process, which is why its spectrum is <u>not</u> symmetric about zero.

cross-spectral density is given by

$$g_{yz}(e^{-iw}) = h_{12}(e^{-iw})h_{34}(e^{+iw})\cdot g_x(e^{-iw})$$

$$= 0 \quad .$$

Therefore, the y and z processes are orthogonal at all lags. In particular we have (applying (89))

$$E\{y_{w_1w_2}(0)\cdot z_{w_3w_4}{*}(0)\} = E\{[G(w_1) - G(w_w)][G^*(w_3) - G^*(w_4)]\} = 0 \quad .$$

Thus, G(w) is a process with orthogonal increments.

Finally we have, since $Ex_t=0$, applying (89)

$$E\{G(w_1) - G(w_2)\} = \sum_{t=-\infty}^{\infty} \frac{e^{-iw_1t} - e^{-iw_2t}}{-it} Ex_t = 0 \quad .$$

Collecting these properties, we have

(a) If $\pi \geq w_1 > w_2 \geq -\pi$,

$$E\{|G(w_1) - G(w_2)|^2\} = 2\pi \int_{w_2}^{w_1} g_x(e^{-iw})dw$$

(b) If $\pi \geq w_1 > w_2 \geq w_3 > w_4 > -\pi$,

$$E\{[G(w_1) - G(w_2)][G^*(w_3) - G^*(w_4)]\} = 0 \quad .$$

(c) $$E[G(w_1) - G(w_2)] = 0 \quad .$$

In summary, the Cramér representation theorem assures us that for every covariance stationary stochastic process $\{x_t\}_{t=-\infty}^{\infty}$ with mean zero, there exists a related complex-valued stochastic process G(w), $w\epsilon[-\pi,\pi]$ such that

$$x_t = \frac{1}{2\pi} \int e^{+iwt} dG(w) \quad .$$

It is properties (a) and (b) of the stochastic process G(w) that motivate the interpretation of the spectrum as representing an orthogonal decomposition by frequency of the variance of x_t.

Vector Stochastic Difference Equations

Let x_t be an (nx1) - vector wide-sense stationary stochastic process that is governed by the matrix difference equation

$$(91) \qquad C(L)x_t = \varepsilon_t$$

where ε_t is now an (nx1)-vector of white noises with means of zero and contemporaneous covariance matrix $E\varepsilon_t\varepsilon_t' = V$, an (nxn) matrix. We assume $E\varepsilon_t\varepsilon_{t-s}' = 0_{(nxn)}$ for all $s \neq 0$. In (91), C(L) is an (nxn) matrix of (finite order) polynomials in the lag operator L:

$$C(L) = \begin{bmatrix} C_{11}(L) & C_{12}(L) & \cdots & C_{1n}(L) \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ C_{n1}(L) & \cdots\cdots & & C_{nn}(L) \end{bmatrix}$$

where each $C_{ij}(L)$ is a finite order polynomial in the lag operator.

We assume that the matrix C(L) has an inverse under convolution $C(L)^{-1} \equiv B(L)$; $C(L)^{-1}$ is defined as the matrix which satisfies

$$C(L)^{-1}C(L) = I_{(nxn)}$$

where $I_{(nxn)}$ is the (nxn) identity matrix. If it exists, $C(L)^{-1}$ can be found as follows. Evaluate the matrix z transform C(z) at $z=e^{-iw}$ to get

$C(e^{-iw})$.

Then invert $C(e^{-iw})$, frequency by frequency, to get $C(e^{-iw})^{-1}$. Finally, the matrix coefficients $C(L)^{-1}=B(L)=\sum_{j=0}^{n} B_j L^j$, B_j being an (nxn) matrix, can be found from the inversion formula

$$B_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} C(e^{-iw})^{-1} e^{iwj} dw,$$

where by integrating a matrix we mean to denote element-by-element integration.

The solution of (91) is found by premultiplying (91) by B(L) to obtain

(92) $$x_t = B(L)\varepsilon_t.$$

The vector stochastic difference equation

$$C(L)x_t = \varepsilon_t,$$

is said to be an <u>autoregressive representation</u> for the vector process x_t. The solution

$$x_t = B(L)\varepsilon_t$$

is said to be a vector <u>moving average representation</u> for the process x_t. The cross spectral density matrix of the (nx1) x_t process (which has the cross spectrum between the i^{th} and j^{th} components of x in the $(i,j)^{th}$ position) is given by

(93) $$g_{xx}(e^{-iw}) = B(e^{-iw})VB(e^{+iw})'$$

where ' denotes transposition. Formula (92) is analogous to the univariate equation (35), and can be derived by comparable methods. Alternatively, one can proceed by using the methods of Section ___, and taking the Fourier transform of (92) to get

$$x(w) = B(e^{-iw})\varepsilon(w).$$

Multiplying each side by $T\overline{x(w)}'$ gives

$$Tx(w)\overline{x(w)}' = B(e^{-iw})T\varepsilon(w)\overline{\varepsilon(w)}'B(e^{+iw})'.$$

Taking expected values, and noting that $ET\varepsilon(w)\overline{\varepsilon(w)}'=V$ then gives equation (93).

Equation (93) is very compact formula for calculating the cross spectra of the (nx1) x_t process as a function of the fundamental parameters, the covariance matrix V and the coefficients in C(L) (or B(L)). Equation (91) is quite a general representation and is flexible enough to incorporate <u>exogenous variables</u> and serially correlated noises.

In equation (91), a variable x_{it} is said to be exogenous if $C_{ij}(L)=0$ for all j not equal to i. This means that the row of equation (91) corresponding to x_{it} becomes

$$C_{ii}(L)x_{it} = \varepsilon_{it},$$

so that x_{it} is governed by only its own past interacting with the random shock ε_{it}. In this sense, the evolution of x_{it} is not affected by interactions with other variables in x_t. This is <u>not</u> to say, however, that x_{it} is uncorrelated with other components of x_t, since ε_{it} can be correlated contemporaneously with other ε's (that is, V need not be diagonal). The definition of exogeneity given here turns out to be precisely the one used by econometricians in a time series context (see Section ___ below).

Serially correlated errors can be incorporated by suitably redefining the errors as components of x_t, and then modeling them as exogenous processes that affect but aren't affected by other components of x_t.

An Example

As an example, consider the following system of stochastic difference equations:

$$p_t - {}_tp^*_{t-1} = .2(y_t - \bar{y}_t) + \varepsilon_{1t} \qquad \text{(Phillips curve)}$$

$$\bar{y}_t = .75n_t + .25k_t \qquad \text{(capacity output equation)}$$

$$k_t - k_{t-1} = .05(y_{t-1} - k_{t-1}) - .5(r_{t-1} - ({}_tp^*_{t-1} - p_{t-1})) + \varepsilon_{2t}$$

(Investment schedule)

$${}_tp^*_{t-1} = 1.5\, p_{t-1} - .5p_{t-2}$$

(formation of price expectations)

$$m_t - p_t = y_t - 10r_t \qquad \text{(LM schedule)}$$

$$y_t - k_t = -6.0(r_t - ({}_{t+1}p^*_t - p_t)) + u_{3t} \qquad \text{(IS schedule)}$$

$$m_t = .75m_{t-1} + \varepsilon_{3t} \qquad \text{(exogenous money supply process)}$$

$$u_{3t} = .75\, u_{3t-1} + \varepsilon_{4t} \qquad \text{(process for IS curve shock)}$$

$$n_t = 0 \qquad \text{(a detrending device)}$$

The $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)$ process is serially uncorrelated with contemporaneous covariance matrix

$$V = \begin{pmatrix} .0001 & 0 & 0 & 0 \\ 0 & .000049 & 0 & 0 \\ 0 & 0 & .01 & 0 \\ 0 & 0 & 0 & .09 \end{pmatrix}$$

Have y_t is the log of real GNP, p_t the log of the GNP deflator, k_t the log of the capital stock, m_t the log of the money supply, n_t the log of the labor supply, ${}_tp^*_{t-1}$ the public's expectation of p_t formed as of time (t-1), and r_t the level of the interest rate. The system is essentially a stochastic version of the dynamic Keynesian model that we analyzed earlier. To induce stationarity in the processes, we have set n_t=0, which has the effect of requiring that our calculations be regarded as recording the spectral densities of the variates expressed as deviations from trends.

Table 1 plots the spectral densities of several variables, while Tables 2 and 3 record various coherence and gain or amplitude of cross spectrum which were calculated using the formulas of the preceding section.

Notice that the spectral density of real GNP has a peak in the vicinity of a 38 period cycle, which with quarterly data would amount to about a nine year cycle, about the length of NBER major cycles.

The model generates a "Gibson paradox," which is to say there is high coherence between the price level and the interest rate at low frequencies.

Notice that the gain of the log of real GNP against the log of money is zero at zero frequency, while the gain of the log of price against the log of money is unity at zero frequency. As we saw earlier, the gain at zero frequency equals the sum of the distributed lag weights, so that these results are consistent with the "classical" long-run character of the present model. In the long run, prices respond proportionately to the money supply while real GNP shows no long-run reponse.

For use of these techniques to analyze actual estimated econometric models, the reader is directed to Howrey [ ] and Chow [ ].

A Compact Notation

It is always possible to write an mth order difference equation in terms of a vector first order system. For example, consider the bivariate system

$$\begin{aligned} x_{1,t+1} &= \alpha_1 x_{1,t} + \dots + \alpha_m x_{1,t-m+1} + \alpha_{m+1} x_{2,t} + \dots \\ &\quad + \alpha_{2m} x_{2,t-m+1} + \varepsilon_{1,t+1} \end{aligned}$$

(94)

$$\begin{aligned} x_{2,t+1} &= \beta_1 x_{1,t} + \dots + \beta_m x_{1,t-m+1} + \beta_{m+1} x_{2,t} + \dots \\ &\quad + \beta_{2m} x_{2,t-m+1} + \varepsilon_{2,t+1} \end{aligned}$$

where $(\varepsilon_{1,t+1}, \varepsilon_{2,t+1})$ are two serially uncorrelated white noise processes. Equations (57) can be written as

(95) $$x_{t+1} = Ax_t + \varepsilon_{t+1}$$

where

$$A = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_m & \alpha_{m+1} & \cdots & \alpha_{2m} \\ 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & & & & \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ \beta_1 & \beta_2 & \cdots & \beta_m & \beta_{m+1} & \cdots & \beta_{2m} \\ 0 & 0 & & 0 & 1 & 0 \cdots & 0 \\ \vdots & & & & & & \\ 0 & 0 & & 0 & 0 \cdots 0 & 1 & 0 \end{pmatrix} \leftarrow (m+1)^{st} \text{ row}$$

$\uparrow (m+1)^{st}$ column

$$x_{t+1} = \begin{pmatrix} x_{1,t+1} \\ x_{1,t} \\ \vdots \\ x_{1,t-m+2} \\ x_{2,t+1} \\ x_{2,t} \\ \vdots \\ x_{2,t-m+2} \end{pmatrix} \qquad \varepsilon_{t+1} = \begin{pmatrix} \varepsilon_{1,t+1} \\ 0 \\ 0 \\ \vdots \\ 0 \\ \varepsilon_{2,t+1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow (m+1)^{th} \text{ row}$$

The solution of the vector difference equation (95) can be written

(96) $$x_{t+\tau} = A^{\tau}x(t) + \varepsilon(t+\tau) + A\varepsilon(t+\tau-1) + \ldots + A^{\tau-1}\varepsilon(t+1) \quad .$$

Since $\quad E\varepsilon(t+\tau)x(t)' = 0_{2mx2m} \qquad$ for all $\tau \geq 1$,

multiplying the solution (96) through by $x(t)'$ and taking expected values gives the matrix Yule-Walker equation

$$Ex_{t+\tau}x_t{}' = A^{\tau}Ex_tx_t{}' \qquad \tau > 1$$

or

$$C_x(\tau) = A^{\tau}C_x(0) \qquad \tau > 1 \tag{97}$$

where $C_x(\tau)=Ex(t+\tau)x(t)'$. As before, we have the result that the covariogram (this time the matrix covariogram) obeys the deterministic part of the difference equation with initial conditions given by the lagged covariances that are in $C_x(0)$.

Using the compact notation (95), it is straightforward to show that the cross spectral density matrix of the vector x process is given by

$$\textstyle\sum_x(e^{-iw}) = (e^{iw}I-A)^{-1}V(Ie^{-iw}-A')^{-1} \tag{98}$$

where $V=E\varepsilon_t\varepsilon_t{}'$, and where it is assumed that the process is stationary, which requires that the eigenvalues of A have absolute values less than unity.

Assuming that the eigenvalues of A are distinct, it is possible to represent A in the form

$$A = P\Lambda P^{-1}$$

where the columns of P are the eigenvectors of A while Λ is the diagonal matrix whose diagonal entries are the eigenvalues of A. Then we have

$$A^{\tau} = P\Lambda^{\tau}P^{-1},$$

so that the solution (97) can be written

$$C_x(\tau) = P\Lambda^{\tau} P^{-1} C_x(0).$$

This expression shows how the eigenvalues of A govern the behavior of the solution. It also illustrates how increasing the number of variables in the system or increasing the number of lags in any particular equation, increases the order of the A matrix, and thereby contributes to the potential for generating complicating covariograms. Reference to this point can be used to show, for example, that while a one-variable, first-order difference equation can't deliver a covariogram with damped oscillations of period greater than two periods (the periodicity if the single root is negative), a multivariate, first-order (i.e., single lag) system can have complex roots and may, therefore, generate oscillatory covariograms.

Optimal Prediction: Compact Notation

Using the fact that ε_t in (95) is a serially uncorrelated vector process, it is straightforward to deduce from (96) that the projection of $x_{t+\tau}$ against x_t is given by

$$\text{(99)} \qquad P[x_{t+\tau} | x_t] = A^{\tau} x_t.$$

This is a compact formula for linear-least squares predictors of a vector governed by a finite order stochastic difference equation.

Solution Concepts

We have now encountered two concepts of a solution for a system of stochastic difference equations. The first concept, one that agrees with the concept of the solution of a nonstochastic difference equation, is given in compact, form by equation (92): for a given sample path of ε_t's, the particular sample path of x_t's that solves (91) is called the solution. This concept takes the sample path of the random variable ε_t (often called a "realization" of the ε_t process) as given, and proceeds to solve the difference equation as if it were a nonstochastic one with a forcing function given by the particular sample path for the ε_t's.

The second definition of a solution views the input to the solution as being (characteristics of) the probability distribution of the exogenous variables and random disturbances, while the output is (characteristics of) the probability distribution of the endogenous variables. In our case, we are concerned with the first and second moments of the variables in question. This solution concept is summarized compactly in our equation (93),

$$g_{xx}(e^{-iw}) = B(e^{-iw})VB(e^{iw})',$$

in which the cross-spectral density matrix of the (nx1) stochastic process x_t is determined as a function of V, the contemporaneous covariance matrix of the ε's, and the coeffficients that are impounded in $B(e^{-iw})$. Of course, the cross spectral density matrix $g_{xx}(e^{-iw})$ determines all of the covariograms that we're interested in.

It is this second solution concept which determines the moments of the endogenous variables in terms of the moments of the "input" variables that underlies the theory of macroeconometric policy evaluation.

The Relationship between Granger-Wiener Causality and Econometric Exogeneity

Let $\begin{pmatrix} x_t \\ y_t \end{pmatrix}$ be a bivariate, jointly covariance stationary stochastic process. Suppose that $\begin{pmatrix} x_t \\ y_t \end{pmatrix}$ is a strictly linearly indeterministic process with mean zero. Under these conditions, the bivariate version of Wold's theorem states that there exists a moving average representation of the (x_t, y_t) process

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} c^{11}(L) & c^{22}(L) \\ c^{21}(L) & c^{22}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}$$

where $c^{ij}(L) = \sum_{k=0}^{\infty} c_k^{ij} L^k$ are square summable polynomials in the lag operator L that are one-sided in nonnegative powers of L; ε_t and u_t are serially uncorrelated processes with $Eu_t\varepsilon_s=0$ for all t,s; $E\varepsilon_t^{\,2}=\sigma_\varepsilon^{\,2}$, $Eu_t^{\,2}=\sigma_u^{\,2}$; and where the one-step ahead prediction errors are given by $x_t - P[x_t|x_{t-1},\ldots,y_{t-1},\ldots] = c_0^{11}\varepsilon_t + c_0^{12}u_t$, $y_t - P[y_t|x_{t-1},\ldots,y_{t-1},\ldots] = c_0^{21}\varepsilon_t + c_0^{22}u_t$, i.e. ε and u are "jointly fundamental for x and y." Wold's theorem establishes the sense in which a vector moving average is a general representation for an indeterministic covariance stationary vector process. The theorem can be proved by pursuing the same kind of projection arguments used in proving the univariate version of the theorem. Below, we will show how to construct a Wold representation from knowledge of the covariograms of x and y and their cross-covariogram.

We now make the further assumption that the (x_t, y_t) process has an autoregressive representation. In particular, think of constructing a sequence of projections

$$(100) \qquad \begin{pmatrix} x_t \\ y_t \end{pmatrix} = F_1^n \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \ldots + F_n^n \begin{pmatrix} x_{t-n} \\ y_{t-n} \end{pmatrix} + \begin{pmatrix} a_{xt}^n \\ a_{yt}^n \end{pmatrix}$$

where $F_1^n, \ldots, F_n^n$ are (2x2) matrices of least squares coefficients and we have the orthgonality conditions

$$E \begin{pmatrix} x_{t-j} \\ y_{t-j} \end{pmatrix} [a_{xt}^n a_{yt}^n] = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

for $j=1,\ldots,n$. We assume that as $n \to \infty$, the F_j^n's converge to F_j for each j. This is the assumption that (x_t, y_t) possesses an autoregressive representation and is stronger than the conditions required for (x_t, y_t) to have a vector moving average representation. We can write the autoregressive representation for (x_t, y_t) as

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \sum_{j=1}^{\infty} F_j \begin{pmatrix} x_{t-j} \\ y_{t-j} \end{pmatrix} + \begin{pmatrix} a_{xt} \\ a_{yt} \end{pmatrix}$$

$$= F(L) \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} a_{xt} \\ a_{yt} \end{pmatrix}, \qquad F(L) = \sum_{j=1}^{\infty} F_j L^{j-1}$$

where the random variables (a_{xt}, a_{yt}) obey the least squares orthogonality conditions

$$E \begin{pmatrix} x_{t-j} \\ y_{t-j} \end{pmatrix} [a_{xt} a_{yt}] = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

for all $j \geq 1$. The random variables (a_{xt}, a_{yt}) are the one-step-ahead errors in predicting (x_t, y_t) from all past values of x and y.

Now consider obtaining the following representation for the (x_t, y_t) process:

(101)
$$A(L) \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

or

$$(A_0 - A_1 L - A_2 L^2 \ldots) \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

where A_j is a (2x2) matrix for each j, where A_0 is chosen to be lower triangular and

$$\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

are pairwise orthogonal processes (at all lags) that are serially uncorrelated. Can we be sure that such a representation can be arrived at, in particular one with A_0 being lower triangular and ε and u being orthogonal processes?

The answer is in general yes,* as the following argument suggests. Think of projecting x_t against all <u>lagged</u> x's and <u>lagged</u> y's. This gives the first row of A(L) and gives a least squares residual process ε_t that is by construction orthogonal to all lagged y's and all lagged x's. Next project y_t against <u>current</u> and lagged x's and all lagged y's. This gives the second row of A(L) and delivers a disturbance process u_t that is by construction orthogonal to <u>current</u> and <u>lagged</u> x's and lagged y's. This procedure produces an A_0 that is lower triangular as required. Further, notice that since ε_t is orthogonal to all lagged x's and y's, and since the representation (101) that we have achieved permits lagged ε's and u's to be expressed as linear combinations of lagged x's and y's, it follows that ε_t is orthogonal to lagged u's and ε's. A similar argument shows that u_t is orthogonal to lagged u's and ε's. Finally, since by construction u_t is orthogonal to current and lagged x's and lagged y's, and since ε_t is by definition a linear combination of

---

*We have remarked earlier that the vector moving average representation of a vector process z_t in terms of the vector noise n_t

$$z_t = C(L)n_t ,$$

where the components of n_t are white noises that are mutually orthogonal at all lags, is a very general representation. An autoregressive representation for z_t can be obtained by inverting the preceding equation to get

$$A(L)z_t = n_t$$

where $A(L)=C(L)^{-1}$, which is to say

$$A(e^{-iw}) = C(e^{-iw})^{-1}$$

for each w between $-\pi$ and π. The autoregressive representation exists provided that $C(e^{-iw})$ is invertible at each frequency between $-\pi$ and π. This condition is a restriction but is one that can usually be assumed in applied work. (For an example of a $C(e^{-iw})$ that violates the condition, consider the univariate example $C(e^{-iw})=1-e^{-iw}$--the transform of the first difference operator (1-L)--which equals zero at w=0 and so is not invertible there.)

current and lagged x's and lagged y's, it follows that u_t and ε_t are orthogonal contemporaneously.

To check that he understands this construction, the reader is invited to verify that it would also be possible to choose A_0 to be upper triangular with a new and generally different error process

$$\begin{pmatrix} u'_t \\ \varepsilon'_t \end{pmatrix}$$

that satisfies the same conditions on second moments that the $\left[\begin{smallmatrix} u \\ \varepsilon \end{smallmatrix}\right]$ process satisfies.

To get (101) in a form that is useful for studying prediction problems, premultiply (101) by A_0^{-1} to get*

$$A_0^{-1} A(L) \begin{pmatrix} x_t \\ y_t \end{pmatrix} = A_0^{-1} \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}$$

---

*Notice that (102) is identical with (100) for $n=\infty$, so that we must have

$$F_j = A_0^{-1} A_j$$

$$\begin{pmatrix} a_{xt} \\ a_{yt} \end{pmatrix} = A_0^{-1} \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix} .$$

Notice that (a_{xt}, a_{yt}) are by the orthogonality conditions serially uncorrelated and uncorrelated with one another at all nonzero lags.

or

$$(102)\qquad I\begin{pmatrix} x_t \\ y_t \end{pmatrix} = A_0^{-1}[A_1L + A_2L^2 + \ldots]\begin{pmatrix} x_t \\ y_t \end{pmatrix} + A_0^{-1}\begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}$$

$$= A_0^{-1}H(L)\begin{pmatrix} x_t \\ y_t \end{pmatrix} + A_0^{-1}\begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}$$

where $H(L) = A_1L + A_2L^2 + \ldots$. The linear least squares prediction of the $\begin{pmatrix} x_t \\ y_t \end{pmatrix}$ process based on all lagged x's and all lagged y's (call it $P_{t-1}[{x_t \atop y_t}]$) from (102) is then

$$(103)\qquad P_{t-1}\begin{pmatrix} x_t \\ y_t \end{pmatrix} = A_0^{-1}H(L)\begin{pmatrix} x_t \\ y_t \end{pmatrix} = F(L)L\begin{pmatrix} x_t \\ y_t \end{pmatrix} \quad ,$$

since by construction

$$P_{t-1}\begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix} = 0 \; .$$

The one-step ahead prediction errors in predicting the $[{x \atop y}]$ process are given by

$$A_0^{-1}\begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix} \quad .$$

Thus x-prediction errors and y-prediction errors are contemporaneously correlated so long as A_0 is not diagonal. Notice that since A_0 is lower triangular, so is A_0^{-1}, so that ε_t is the one-step ahead prediction error in predicting x from past x's and y's, which is what should be expected given the way the ε_t process was constructed above.

If $A_0^{-1}A(L)$ is lower triangular (that is, the matrix coefficient is lower triangular for <u>each</u> power of L), then given lagged x's, lagged y's don't help predict current x. That is, if $A_0^{-1}A(L)$ is lower triangular, and, therefore, so is $A_0^{-1}H(L)$, then $P_{t-1}x_t$ involves only lagged x's, lagged y's all bearing zero regression coefficients. In the language of Norbert Wiener and C.W.J. Granger, y is said to <u>cause</u> x if given past x's, past y's help predict current x. Thus, the lower triangularity of $A_0^{-1}A(L)$ is equivalent with y's <u>failing</u> to cause x, in the Wiener-Granger sense.

We now claim the following: $A_0^{-1}A(L)$ is lower triangular, if and only if $A(L)^{-1}$ is lower triangular. To show this, suppose first that $A_0^{-1}A(L)$ is lower triangular. Then note

$$A(L)^{-1} = A(L)^{-1}A_0A_0^{-1}$$

But we know that $A(L)^{-1}A_0$, being the inverse of $A_0^{-1}A(L)$, is lower triangular, as is A_0^{-1}. Noting that the product of two lower triangular matrices is also lower triangular then proves that $A(L)^{-1}$ is lower triangular.*

---

* To make the argument in terms of ordinary matrices, write

$$A(e^{-iw})^{-1} = A(e^{-iw})^{-1}A_0A_0^{-1}$$

and note that $A(e^{-iw})^{-1}A_0$ is the inverse of the lower triangular matrix $A_0^{-1}A(e^{-iw})$ at each frequency and so is lower triangular. It follows that $A(e^{-iw})^{-1}$ is lower triangular (at each frequency) being the product of two lower triangular matrices. It then follows that

$$A_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(e^{-iwj})^{-1}e^{iwj}dw$$

is lower triangular for j=0,1,2,... .

Now suppose that $A(L)^{-1}$ is lower triangular. Since A_0 is lower triangular, it follows that $A_0^{-1} A(L)$ is lower triangular. So we have proved that $A_0^{-1} A(L)$ is lower triangular if and only if $A(L)^{-1}$ is lower triangular.

This establishes that if $A_0^{-1}A(L)$ is lower triangular, then (101) can be "inverted" to yield the vector moving average representation

(103) $$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = C(L) \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}$$

where $A(L)^{-1} = C(L) = C_0 + C_1L + C_2L^2 + \ldots$, C_j being a 2x2 matrix, and where C(L) is lower triangular. Recall the extensive orthogonality conditions satisfied by ε and u: the ε and u processes are orthogonal at all lags, even contemporaneously.* Conversely, suppose that a moving average

---

* Assuming that things have been normalized so that ε and u have unit variances, the spectral density matrix of the (x,y) process satisfying (66) is, as we have seen,

$$S(e^{-iw}) = C(e^{-iw})IC(e^{-iw})'$$

where the prime now denotes both complex conjugation and transposition. Now let be a (2x2) unitary matrix, i.e., a (2x2) matrix satisfying $UU'=U'U=I$ where here the ' again denotes complex conjugation and transposition. Then note that $S(e^{-iw})$ can also be represented

$$S(e^{-iw}) = C(e^{-iw})\ U\ I\ U'C(e^{-iw})'$$

$$S(e^{-iw}) = [C(e^{-iw})U]I[C(e^{-iw})U]'$$

$$S(e^{-iw}) = D(e^{-iw})ID(e^{-iw})'$$

where $D(e^{-iw})=C(e^{-iw})U$. Thus, we have produced a new moving average representation, one with contemporaneously orthogonal disturbances. This proves that a moving average representation is unique only up to multiplication by a unitary matrix. Notice that multiplication of $C(e^{-iw})$ by U will, in general, destroy the lower triangularity of $C(e^{-iw})$ if C originally has this property.

representation of the lower triangular form (103) exists with ε_t and u_t being serially uncorrelated processes with $E\varepsilon_t u_s=0$ for all t and s. Then assuming that $C(L)^{-1}$ exists and equals A(L) gives a representation

$$C(L)^{-1} \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}$$

or

$$A(L) \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}$$

where A(L) is lower triangular and one-sided on the present and past. It follows, then, that y fails to Granger-cause x.

We have now established Sims's important theorem 1, which states:

> Let (x_t, y_t) be a jointly covariance stationary, strictly indeterministic process with mean zero. Then $\{y_t\}$ fails to Granger-cause $\{x_t\}$ if and only if there exists a vector moving average representation
>
> $$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} C^{11}(L) & 0 \\ C^{21}(L) & C^{22}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}$$
>
> where ε_t and u_t are serially uncorrelated processes with means zero and $E\varepsilon_t u_s = 0$ for all t and s, and where the one-step ahead prediction errors $(x_t - P[x_t | x_{t-1}, \ldots, y_{t-1} \ldots])$ and $(y_t - P[y_t | x_{t-1}, \ldots, y_{t-1}, \ldots])$ are each linear combinations of ε_t and u_t.

We are now in a position to state a second theorem of Sims that characterizes the relationship between the concept of strict econometric exogeneity and Granger's concept of causality. Sims's theorem is this:

> y_t can be expressed as a distributed lag of current and past x's (with no future x's) with a disturbance process that is orthogonal to past, present, and future x's if and only if y does not Granger cause x.

The condition that y can be expressed as a one-sided distributed lag of x with disturbance process that is orthogonal at all lags to the x process is known as the strict econometric exogeneity of x with respect to y. In applied work it is important to test for this condition, since the condition is required if various estimators are to have good properties. It is interesting that engineers have long called a relationship in which y is a one-sided (on the present and past) distributed lag of x a "causal" relationship, and that this long-standing use of the word cause should happen to coincide with the failure of y to cause x in the Wiener-Granger sense.

First we prove that y's not Granger causing x implies that y can be expressed as a one-sided distributed lag of x with a disturbance process orthogonal to x at all lags. The lack of Granger causality from y to x is equivalent with $A_0^{-1}A(L)$ being lower triangular. As we have seen, this implies that C(L) in (103) is lower triangular, so that

(104) $$x_t = C^{11}(L)\varepsilon_t$$

(105) $$y_t = C^{21}(L)\varepsilon_t + C^{22}(L)u_t$$

where all polynomials in L involve only nonnegative powers of L.

Inverting (104) and substituting into (105) gives*

$$y_t = C^{21}(L)C^{11}(L)^{-1}x_t + C^{22}(L)u_t$$

which expresses y_t as a one-sided distributed lag of x (no negative powers of L enter) with a disturbance process u_t that is orthogonal to ε_t and therefore to x_t at all lags. This proves half of the theorem.

To prove the other half, one would start with a one-sided lag distribution and a moving average representation for x_t

$$y_t = h(L)x_t + \eta_t$$

$$x_t = a(L)\varepsilon_t$$

where by hypothesis η is orthogonal to ε and therefore to x at all lags. Then by finding the moving average representation for η_t, say

$$\eta_t = m(L)u_t$$

where $Eu_t\varepsilon_s=0$ for all t,s, one gets the lower triangular vector moving average representation

$$y_t = h(L)a(L)\varepsilon_t + m(L)u_t$$

$$x_t = a(L)\varepsilon_t$$

or

---

* In assuming that $\begin{pmatrix} x_t \\ y_t \end{pmatrix}$ has an autoregressive representation, we have in effect assumed that $C^{11}(L)$ has an inverse that is one-sided in nonnegative powers of L.

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = C(L) \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

where C(L) is lower triangular. Assuming that $C(L)^{-1}$ exists then gives

$$C(L)^{-1} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

where $C(L)^{-1}$ is lower triangular and say equal to A(L). Multiplying the above equation, which is in the form of (103), through by A_0^{-1}, which is also lower triangular then gives

$$A_0^{-1}A(L) \begin{bmatrix} x_t \\ y_t \end{bmatrix} = A_0^{-1} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

or

$$[I-A_0^{-1}A_1L-A_0^{-1}A_2L^2-\ldots] \begin{bmatrix} x_t \\ y_t \end{bmatrix} = A_0^{-1} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix},$$

The lower triangularity of the matrices on the left and the orthogonality properties of ε and u establish that in this system y does not Granger cause x, i.e., y does not help predict x given lagged x's. This proves the other half of Sims's theorem 2.

Sims's Application to Money and Income

Economists at the Federal Reserve Bank of St. Louis have computed estimates of one-sided distributed lag regressions of (the log of) nominal income (y_t) against (the log of) money (m_t):

$$(106) \qquad y_t = \sum_{j=0}^{\infty} h_j m_{t-j} + \eta_t \quad ,$$

where $E\eta_t m_{t-j}=0$ for $j=0,1,2,\ldots$. Those economists recommend that the h_j's be taken seriously and be regarded as depicting the response of nominal income to exogenous impulses in the money supply. However, Keynesian economists have tended not to regard the h_j's as good estimates of the response pattern (or "dynamic multipliers") of nominal income to money. Their argument has two parts. First, in the kind of macroeconometric model the Keynesians have in mind, even were it true that money had been made to behave exogenously with respect to nominal income, the "final form" for money income has many additional right-hand-side variables not included in (106), e.g.

$$(107) \qquad y_t = \sum_{j=0}^{\infty} v_j m_{t-j} + \sum_{j=0}^{\infty} w_j z_{t-j} + \varepsilon_t$$

where z_t is a vector of stochastic processes including government tax and expenditures parameters and w_j is a vector conformable to z_t; the error term ε_t is a stationary stochastic process that obeys the orthogonality conditions $E\varepsilon_t m_{t-j} = E\varepsilon_t z_{t-j} = 0$ for $j=0,\pm1,\pm2,\ldots$.

The strong condition that ε must be orthogonal to m and z at all leads and lags is the requirement that m and z be "strictly econometrically exogenous with respect to y" in relation (107). These orthogonality conditions characterize (107) as a "final form" relationship. In (107), the v_j's are the dynamic money multipliers and depict the average response of y_t to a unit impulse in m, holding constant the z's. Applying the law of iterated projections to (107) we obtain

$$P[y_t|m_t,m_{t-1},\ldots] = \sum_{j=0}^{\infty} v_j m_{t-j} + \sum_{k=0}^{\infty} w_k P[z_{t-k}|m_t,m_{t-1},\ldots]$$

Let $$P[z_{t-k}|m_t,m_{t-1},\ldots] = \sum_{j=0}^{\infty} \alpha_{kj} m_{t-j} \quad .$$

Then we have

$$P[y_t|m_t,m_{t-1},\ldots] = \sum_{j=0}^{\infty} v_j m_{t-j} + \sum_{k=0}^{\infty} w_k \sum_{j=0}^{\infty} \alpha_{kj} m_{t-j}$$

or

$$y_t = \sum_{j=0}^{\infty} (v_j + \sum_{k=0}^{\infty} w_k \alpha_{kj}) m_{t-j} + \eta_t$$

where by the orthogonality principle we have $E\eta_t m_{t-j}=0$, $j=0,1,2,\ldots$. Now (108) is identical with (106) so that the population h_j's of (106) obey

$$h_j = v_j + \sum_{k=0}^{\infty} w_k \alpha_{kj} \quad .$$

Therefore, the h_j's in general <u>don't</u> equal the money multipliers, the v_j's. The h_j's are "mongrel" coefficients that do not indicate the typical average response to y to exogenous inpulses in m, everything else being held constant. For this reason, Keynesians would argue, estimating equation (106) is not a good way of estimating the dynamic multipliers, the v_j's.

Now project both sides of (107) against the entire sequence $\{m_{t-j}\}_{j=-\infty}^{\infty}$ to get

$$(109) \qquad y_t = \sum_{j=0}^{\infty} h_j m_{t-j} + \sum_{k=0}^{\infty} w_k \sum_{j=-\infty}^{\infty} \gamma_{kj} m_{t-j} + \xi_t$$

where $E\xi_t \cdot m_{t-j} = 0$ for all j and

$$P\left(z_{t-k} | \{m_{t-j}\}_{j=-\infty}^{\infty}\right) = \sum_{j=-\infty}^{\infty} \gamma_{kj} m_{t-j}$$

where γ_{kj} is a vector of coefficients. We can write (109) as

$$P\left(y_t | \{m_{t-j}\}_{j=-\infty}^{\infty}\right) = \sum_{j=-\infty}^{\infty} d_j m_{t-j}$$

where

$$d_j = h_j + \sum_{k=0}^{\infty} w_k \gamma_{kj} \qquad j \geq 0$$

$$d_j = \sum_{k=0}^{\infty} w_k \gamma_{kj} \qquad j < 0$$

In general, so long as the processes m_t and z_t are correlated (as we had to assume to make the argument that the St. Louis h_j's are mongrel parameters), the γ_{kj}'s and therefore the d_j's will not vanish for some $j<0$. That is because in general future m's will help explain current and past z_t's.* Therefore, so long as the w_k's are not zero in the final form (107), i.e., so long as the z's appear in the final form for y_t, the projection of y_t on current and lagged m's is predicted to be two-sided.

---

* Unless m_t is strictly exogenous with respect to the vector z_t or, equivalently, the vector z_t does not Granger-cause m_t.

For this reason, a test of the null hypothesis that the projection of y_t on the entire $\{x\}$ process is one-sided (i.e. it equals the projection of y_t on current and past x's alone) can be regarded as testing the null hypothesis that the w_k's in (107) <u>are</u> zeroes. But remember that the contention that the w_k's aren't zero is what underlies the Keynesian objection against interpreting the St. Louis equation's h_j's as estimates of the dynamic money multipliers. So computing the two-sided projection

$$(110) \qquad y_t = \sum_{j=-\infty}^{\infty} \delta_j m_{t-j} + \hat{\eta}_t$$

where $E\hat{\eta}_t m_{t-j} = 0$ for all j, and testing the null hypothesis that $\delta_j = 0$ for all $j<0$ provides a means of testing the null hypothesis that the St. Louis equation is "properly specified"--that is, that it is appropriate to set the w_k's equal to zero.

Using post-World War II U.S. data, Sims estimated (110) and implemented the preceding test. He found that he could not reject with high confidence the hypothesis that future m's bear zero coefficients in (110). In general, if the Keynesian objection to the St. Louis equation were correct, in large enough samples one would expect to reject the hypothesis tested by Sims. Sims's particular statistical results have provoked much controversy. Since his tests are subject to usual kinds of type I and type II statistical errors, there is some room for disagreement about how far his results go in confirming using the St. Louis equation to estimate money multipliers. Nevertheless, it should be recognized how much of a contribution Sims made in providing a formal statistical setting in which one could in principle subject to statistical testing the Keynesian claims made against the St. Louis approach. Before Sims's work, those claims were entirely a priori and, though they had been made repeatedly, had never been subjected to any empirical tests.

As it happens, the test implemented by Sims is also useful in discriminating against another hypothesis which has often been advanced to argue that the St. Louis equation (106) is not a legitimate final form (i.e. does not have a disturbance that obeys the requirement that it be orthogonal to past, present, and future m's). The argument is that the money supply fails to be exogenous in (106) because the monetary authority has set m via some sort of feedback rule on lagged y's. For example, it is often asserted that the Federal Reserve "leans against the wind," increasing m faster in a recession, more slowly in a boom. If the Fed behaved this way, it could mean that the projection (106) of y on m partly reflects this feedback from past y to m as well as the effect of m on y. Furthermore, such behavior by the Fed would in general lead us to expect the projection of y on the entire m process to differ from the projection of y on current and past m's, so that the η_t's in (106) would not obey the restrictions $E\eta_t m_{t-s}=0$ for all s; i.e. (106) would not be a final form.

Now Sims's theorems assure us that if the projection of y_t on $\{m_{t-j}\}_{j=-\infty}^{\infty}$ is one-sided on the present and past (as Sims was unable to reject), then there exists a representation (i.e. a model consistent with the data) of the form

$$m_t = C^{11}(L)\varepsilon_t$$

$$y_t = d(L)m_t + C^{22}(L)u_t$$

where $Eu_t\varepsilon_s=0$ for all t,s, and d(L), $C^{11}(L)$, $C^{22}(L)$ are one-sided on the present and past. This representation is one in which there is no feedback from y to m. Thus, Sims's results are consistent with the view that there was no systematic feedback from y to m in the sample period he studied.

Sims's work on money and income was important because it provided a valid framework for testing empirically some often-stated objections to interpreting St. Louis regressions as final form equations.

Bivariate Prediction Formulas

Continue to assume that (x_t,y_t) is a jointly covariance stationary, strictly indeterministic process with a moving average representation

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} C^{11}(L) & C^{12}(L) \\ C^{21}(L) & C^{22}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix} = C(L) \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}$$

where $E\varepsilon_t u_s=0$ for all t,s, $\{\varepsilon_t,u_t\}$ are jointly fundamental for (x_t,y_t), and where $C(L)^{-1}$ exists and is one-sided and convergent in nonnegative powers of L, so that (x_t,y_t) has an autoregressive representation

$$C(L)^{-1} \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}$$

or

$$A(L) \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}$$

where $A(L)=C(L)^{-1}$. Paralleling our calculations in the univariate case, it is easy to deduce that the projection of (x_{t+1},y_{t+1}) against $\{x_t,x_{t-1},\ldots,y_t,y_{t-1},\ldots\}$, call it $P_t\begin{pmatrix}x_{t+1}\\y_{t+1}\end{pmatrix}$, is

$$P_t\begin{pmatrix}x_{t+1}\\y_{t+1}\end{pmatrix} = \left[\frac{C(L)}{L}\right]_+\begin{pmatrix}\varepsilon_t\\u_t\end{pmatrix}$$

$$= \left[\frac{C(L)}{L}\right]_+ A(L)\begin{pmatrix}x_t\\y_t\end{pmatrix} \quad .$$

More generally, we have

$$P_t\begin{pmatrix}x_{t+j}\\y_{t+j}\end{pmatrix} = \left[\frac{C(L)}{L^j}\right]_+ A(L)\begin{pmatrix}x_t\\y_t\end{pmatrix}$$

To take an example, let R_{nt} be the rate of n-period bonds, and assume that (R_{nt},R_{1t}) has moving average representation

(111) $$R_{nt} = \alpha(L)\varepsilon_t + \beta(L)u_t \qquad n>1$$

$$R_{1t} = \gamma(L)\varepsilon_t + \delta(L)u_t$$

where all lag operators are one-sided on the present and past, and

$$R_{nt} - P_{t-1}R_{nt} = \alpha_0\varepsilon_t + \beta_0 u_t$$

$$R_{1t} - P_{t-1}R_{1t} = \gamma_0\delta_t + \delta_0 u_t \quad .$$

The rational expectations theory of the term structure asserts*

---

* Assuming that information used to forecast R_{1t} is confined to current and past R_{1t}'s and R_{nt}'s alone.

$$R_{nt} = \frac{1}{n}\,[R_{1t} + P_t R_{1t+1} + \ldots + P_t R_{1t+n-1}]$$

$$= \frac{1}{n}\,[\gamma(L) + \frac{\gamma(L)}{L} + \ldots + \frac{\gamma(L)}{L^{n-1}}]_+ \,\varepsilon_t$$

$$+ \frac{1}{n}\,[\delta(L) + \frac{\delta(L)}{L} + \ldots + \frac{\delta(L)}{L^{n-1}}]_+ \,u_t$$

or

(112) $$R_{nt} = \frac{1}{n}\,[\left(\frac{1-L^{-n}}{1-L^{-1}}\right)\gamma(L)]_+ \,\varepsilon_t + \frac{1}{n}\,[\left(\frac{1-L^{-n}}{1-L^{-1}}\right)\delta(L)]_+ \,u_t$$

Thus, comparing (111) with (112), it is seen that the rational expectations theory of the term structure imposes the following restrictions across the equations of the moving average representation of the (R_{nt}, R_{1t}) process:

$$\alpha(L) = \frac{1}{n}\,[\left(\frac{1-L^{-n}}{1-L^{-1}}\right)\gamma(L)]_+$$

$$\beta(L) = \frac{1}{n}\,[\left(\frac{1-L^{-n}}{1-L^{-1}}\right)\delta(L)]_+$$

These restrictions embody the content of the theory and are refutable.

Multivariate Prediction Formulas

The results of the last section extend in a natural way to n-dimensional stochastic processes. In particular, the n-variate version of Wold's theorem implies that if $\{y_t\}$ is an n-dimensional, jointly covariance stationary, strictly indeterministic stochastic process with mean zero, it has a moving average representation

(113) $\quad y_t = C(L)\varepsilon_t$

where $C(L) = C_0 + C_1L + \ldots$, C_j being an (nxn) matrix and the C_j's being "square summable," where ε_t is an (n+1) vector stochastic process, where the component ε_{it}'s are serially uncorrelated and mutually orthogonal (at all lags), $E\varepsilon_{it}\varepsilon_{js} = 0$ for all t,s where $i \neq j$; and the ε_{it}'s are "jointly fundamental for y_t," i.e. for each i $(y_{it} - (Py_{it}|y_{t-1}, y_{t-2}, \ldots))$ is a linear combination of ε_{jt}, j=1,...,n. For the process (113) we have the prediction formula

$$E_t t_{t+j} = \left(\frac{C(L)}{L^j}\right)_+ \varepsilon_t$$

where $E_t(x) \equiv Ex|y_t, y_{t-1}, \ldots$. Where $C(L)^{-1}$ exists, so that y_t has a vector autoregressive representation, then we also have the formula

$$E_t y_{t+j} = \left(\frac{C(L)}{L^j}\right)_+ C(L)^{-1} y_t \quad .$$

Solving Rational Expectations Models

This section summarizes the general method that John F. Muth used to solve for stochastic processes that satisfy the restrictions imposed by rational expectations models.

A general linear rational expectations structural model has the form

(114) $$\sum_{j=0}^{\infty} A_j y_{t-j} + \sum_{j=1}^{\infty} B_j E_t y_{t+j} + \eta_t = 0$$

where A_j and B_j are (nxn) matrices, y_t is an (nxl) stochastic process, and η_t is an (nxl)-stochastic process of structural disturbances. Let η_t have a moving average (Wold) representation

$$(115) \qquad \eta_t = \sum_{j=0}^{\infty} F_j \varepsilon_{t-j}$$

$$= F(L)\varepsilon_t$$

where F_j is an (nxn) matrix, $E\varepsilon_{is}\varepsilon_{jt} = 0$ for $i \neq j$ and all t and s; and for all i, $E\varepsilon_{it}\varepsilon_{is} = 0$ for all $t \neq s$; and where ε_t is jointly fundamental for η_t (i.e. for i=1,...,n, we have $\eta_{it} - P[\eta_{it}|\eta_{t-1},\eta_{t-2},\ldots]$ is a linear combination of the ε_{jt}'s, j=1,...,n).

To find a stochastic process that satisfies the stochastic difference equation (114), "guess" that the final solution will have the form

$$(116) \qquad y_t = C(L)\varepsilon_t \quad .$$

Then use the prediction formula

$$P_t y_{t+j} = \left[\frac{C(L)}{L^j}\right]_+ \varepsilon_t \quad .$$

Substituting (114) and (115) into (116) we obtain

$$\left[\sum_{j=0}^{\infty} A_j C(L) L^j\right]\varepsilon_t + \sum_{j=1}^{\infty} B_j \left[\frac{C(L)}{L^j}\right]_+ \varepsilon_t + F(L)\varepsilon_t = 0$$

or

$$A(L)C(L)\varepsilon_t + \left[\sum_{j=1}^{\infty} B_j L^{-j} C(L)\right]_+ \varepsilon_t + F(L)\varepsilon_t = 0$$

or

$$(117) \qquad [A(L)C(L) + [B(L^{-1})C(L)]_+ + F(L)]\varepsilon_t = 0$$

where $A(L) = \sum_{j=0}^{\infty} A_j L^j$, $B(L^{-1}) = \sum_{j=1}^{\infty} B_j L^{-j}$. Equation (117) implies the following equation:

$$(118) \qquad -A(L)^{-1}\{[B(L^{-1})C(L)]_+ + F(L)\} = C(L)$$

which implicitly determines C(L) as a function of the structural parameters A(L), B(L), and F(L). A natural way to solve (118) would be to iterate on it; i.e. notice that given A(L), B(L), and F(L), equation (118) maps one choice of C(L) into another. Start with a guess for C(L) and then use (118) to get a revised guess and hope that the process converges. (In general, there is no guarantee that it will.)

Once C(L) in (118) has been determined, we have determined the stochastic process for y_t. We shall utilize this solution method in the next chapter.

Optimal Filtering Formula

It is convenient to have a formula for the projection of a random variable y_t against current and past values of a covariance stationary, indeterministic random process x_t. We assume that y and x_t have means of zero and are jointly covariance stationary, indeterministic processes. That is, we seek the h_j's that characterize the one-sided projection

(119) $$y_t = \sum_{j=0}^{\infty} h_j x_{t-j} + u_t$$

where $Ex_{t-j}u_t = 0$ for all $j \geq 0$. First, suppose that x_t has the moving average representation

$$x_t = d(L)\varepsilon_t \qquad d(L) = \sum_{j=0}^{\infty} d_j L^j$$

where $\{\varepsilon_t\}$ is a serially uncorrelated process of innovations in x. As an intermediate step,* think of projecting y_t on current and past ε's:

(120) $$y_t = \sum_{j=0}^{\infty} \phi_j \varepsilon_{t-j} + u_t$$

where $Eu_t\varepsilon_{t-j} = 0$ for all $j \geq 0$. We assume that x_t has both a moving average and an autoregressive representation, so that it is easy to see that $\{x_t, x_{t-1}, \ldots\}$ and $\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$ span the same space. For this reason, u_t in (119) equals u_t in (120). Since the ε's form an orthogonal process, we have that the ϕ_j's are the simple least squares coefficients

$$\phi_j = \frac{Ey_t\varepsilon_{t-j}}{E\varepsilon_t^2} = \frac{Ey_t\varepsilon_{t-j}}{\sigma^2}$$

where $\sigma^2 = E\varepsilon_t^2$. Thus we can write

$$\phi(L) = \sum_{j=0}^{\infty} \phi_j L^j$$

---

* This is the method that Wiener used to derive the formula we are after. See Whittle, p. 42.

(121) $\phi(L) = \frac{1}{\sigma^2} [g_{y\varepsilon}(L)]_+$

where $[\;]_+$ again means "ignore negative powers of L" and $g_{y\varepsilon}(L)$ is the cross-covariance generating function

$$g_{y\varepsilon}(L) = \sum_{k=-\infty}^{\infty} E(y_t \varepsilon_{t-k}) L^k \quad .$$

We can relate $g_{y\varepsilon}(L)$ to the cross-covariance generating function $g_{yx}(L)$ as follows:

$$\begin{aligned} g_{yx}(z) &= \sum_k (Ey_t x_{t-k}) z^k \\ &= \sum_k (Ey_t d(L)\varepsilon_{t-k}) z^k \\ &= \sum_k (Ey_t (d_0 \varepsilon_{t-k} + d_1 \varepsilon_{t-k-1} + \ldots)) z^k \\ &= d_0 \sum_k (Ey_t \varepsilon_{t-k}) z^k + d_1 \sum_k (Ey_t \varepsilon_{t-k-1}) z^k \\ &\quad + d_2 \sum_k (Ey_t \varepsilon_{t-k-2}) z^k + \ldots \\ &= d_0 g_{y\varepsilon}(z) + d_1 z^{-1} g_{y\varepsilon}(z) + d_2 z^{-2} g_{y\varepsilon}(z) + \ldots \\ &= d(z^{-1}) g_{y\varepsilon}(z) \quad . \end{aligned}$$

Thus we have

$$g_{y\varepsilon}(z) = \frac{g_{yx}(z)}{d(z^{-1})}$$

Substituting this into (121) we obtain

$$\phi(L) = \frac{1}{\sigma^2}\left[\frac{g_{yx}(L)}{d(L^{-1})}\right]_+$$

So we have

$$y_t = \frac{1}{\sigma^2}\left[\frac{g_{yx}(L)}{d(L^{-1})}\right]_+ \varepsilon_t + u_t$$

(119')

$$y_t = \frac{1}{\sigma^2}\left[\frac{g_{yx}(L)}{d(L^{-1})}\right]_+ \frac{1}{d(L)} x_t + u_t$$

so that in (119) we have

(122)

$$h(L) = \frac{1}{\sigma^2}\left[\frac{g_{yx}(L)}{d(L^{-1})}\right]_+ \frac{1}{d(L)}$$

The classic application of this formula is due to John F. Muth. Suppose that income evolves according to

$$x_t = y_t + \varepsilon_t$$

where $y_t = \rho y_{t-1} + u_t$ $\qquad |\rho|<1$

and where u_t and ε_t are mutually orthogonal at all lags and serially uncorrelated. Here x_t is measured income while y_t is "systematic" or permanent income. The consumer only "sees" $x_t, x_{t-1}, \ldots$ and desires to estimate systematic income y_t by a linear function of $x_t, x_{t-1}, \ldots$. The consumer is assumed to know all the relevant moments. This problem can be solved quickly using formula (122), and the reader is invited to do so. A more tedious method of solution is adopted in section below.

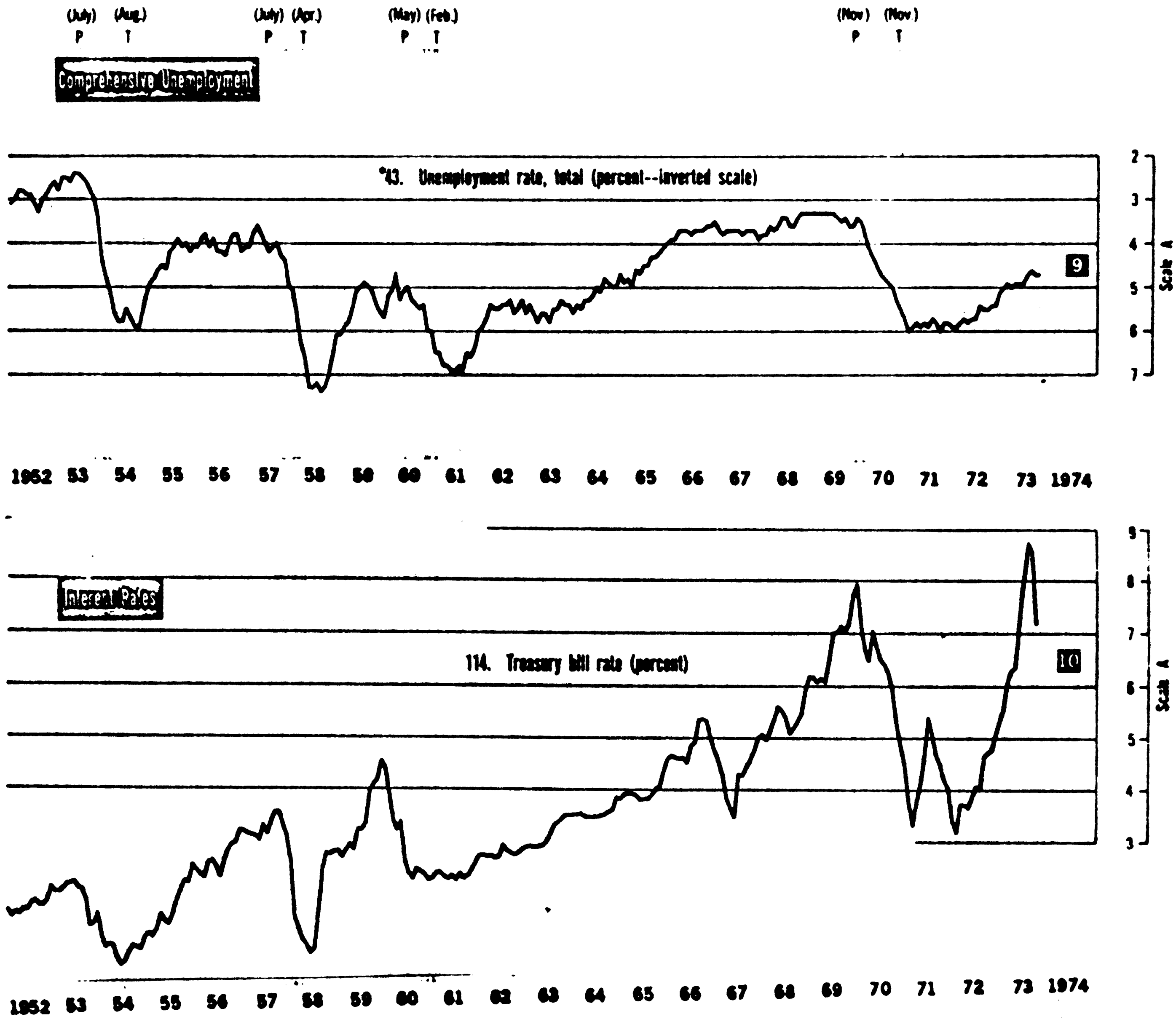

(July) (Aug.)
P T
(July) (Apr.)
P T
(May) (Feb.)
P T
(Nov) (Nov)
P T
Comprehensive Unemployment
43. Unemployment rate, total (percent--inverted scale)
9
Scale A
2
3
4
5
6
7
1952 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 1974
Interest Rates
114. Treasury bill rate (percent)
10
Scale A
9
8
7
6
5
4
3
1952 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 1974

Figure 1

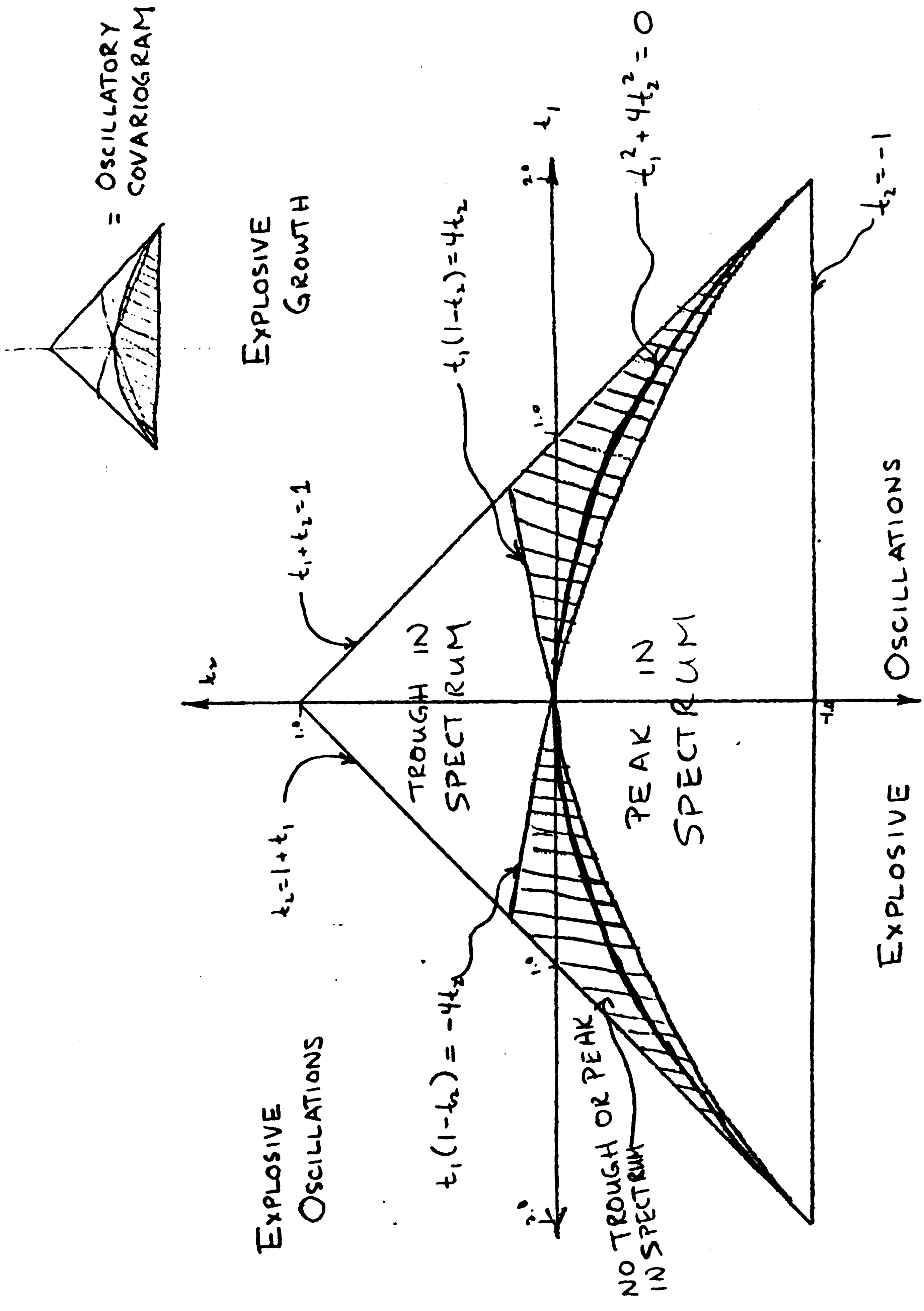

= OSCILLATORY COVARIOGRAM
EXPLOSIVE GROWTH
EXPLOSIVE OSCILLATIONS
TROUGH IN SPECTRUM
PEAK IN SPECTRUM
EXPLOSIVE
OSCILLATIONS
NO TROUGH OR PEAK IN SPECTRUM
$t_1 + t_2 = 1$
$t_2 = 1 + t_1$
$t_1(1-t_2) = 4t_2$
$t_1(1-t_2) = -4t_2$
$t_1^2 + 4t_2^2 = 0$
$t_2 = -1$
t_1
t_2
1.0
2.0
-1.0
-2.0

Figure 2

Stochastic Second Order System
Y(T) = .9Y(T-1) + 0Y(T-2) + EPSILON

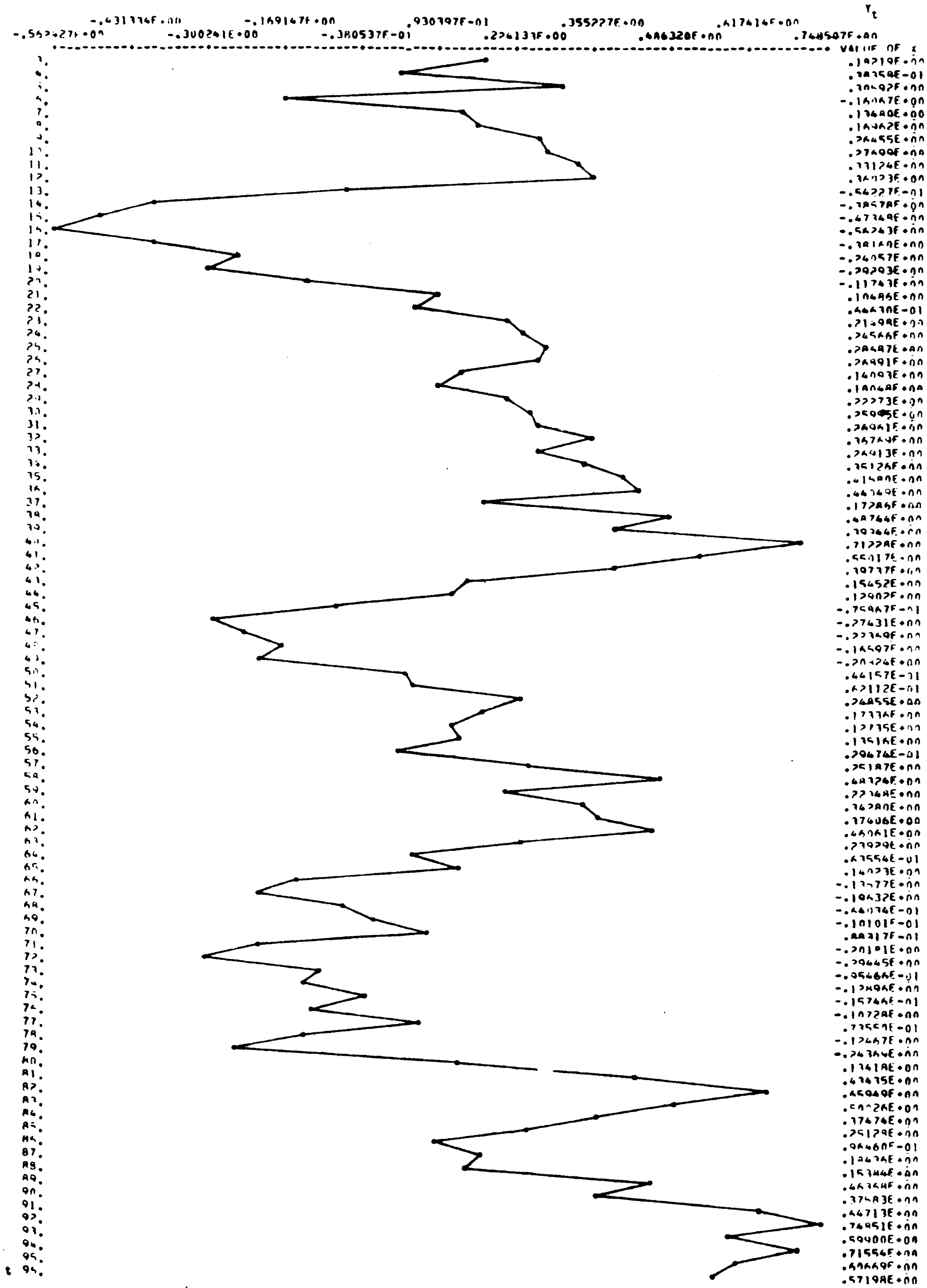


Figure 4a

Stochastic Second Order System
Y(T) = 1.0Y(T-1) + -.5Y(T-2) + EPSILON

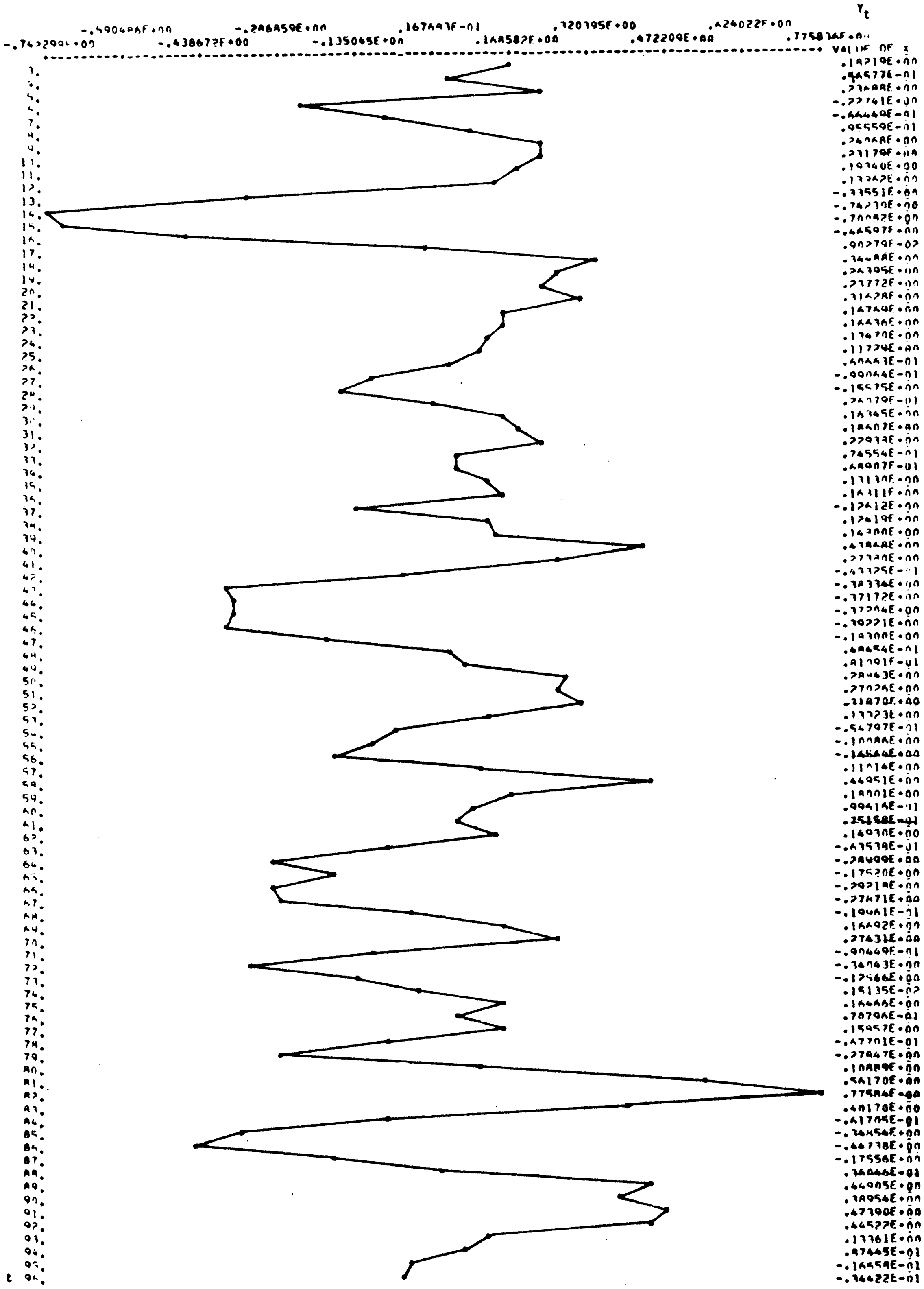


Figure 4b

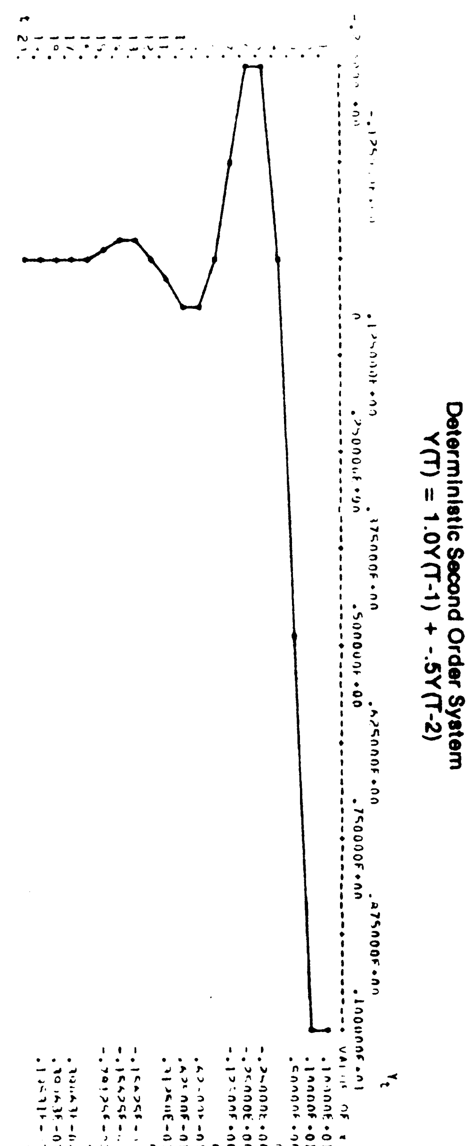
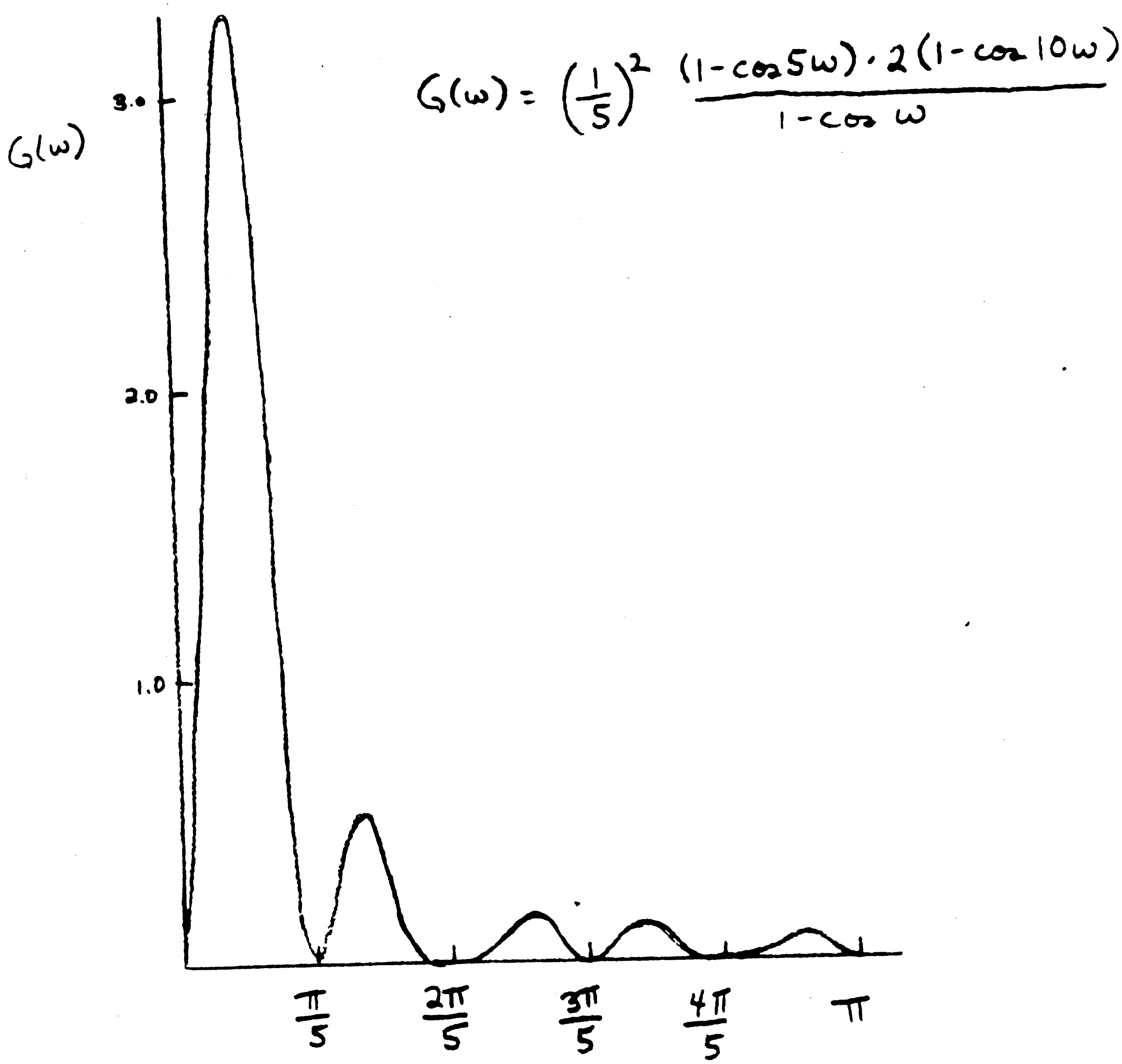


Figure 4c



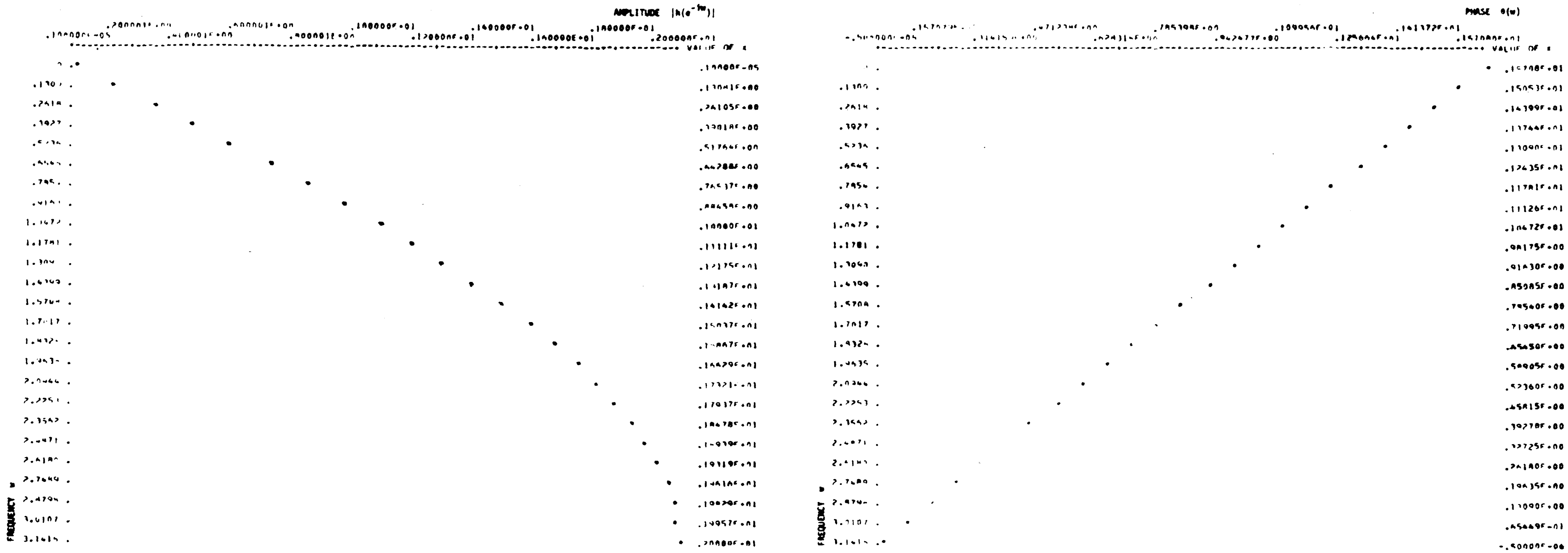
$G(\omega) = \left(\frac{1}{5}\right)^2 \frac{(1-\cos 5\omega)\cdot 2(1-\cos 10\omega)}{1-\cos\omega}$
G(ω)
3.0
2.0
1.0
π/5
2π/5
3π/5
4π/5
π

Figure 5

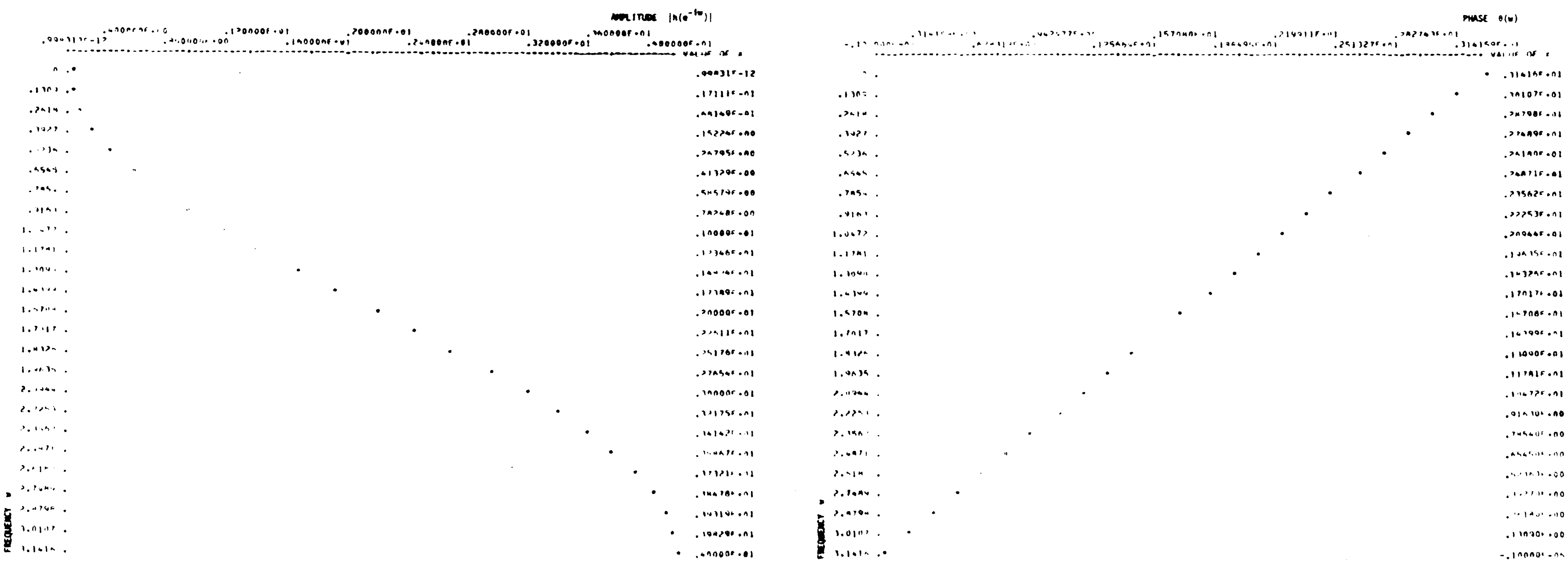
Figure 6

FREQUENCY RESPONSE FUNCTIONS

h(L) = 1 - L

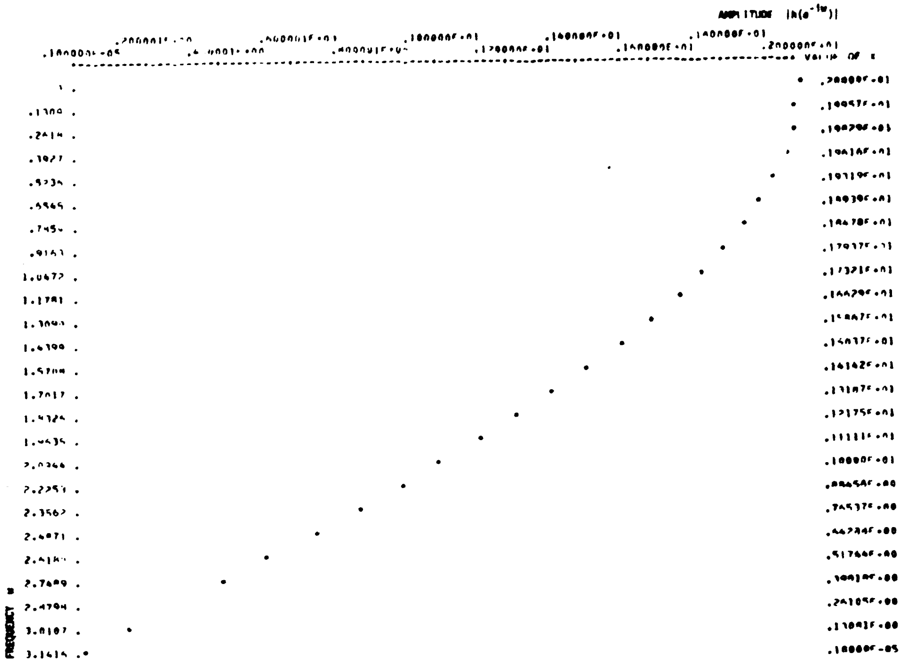


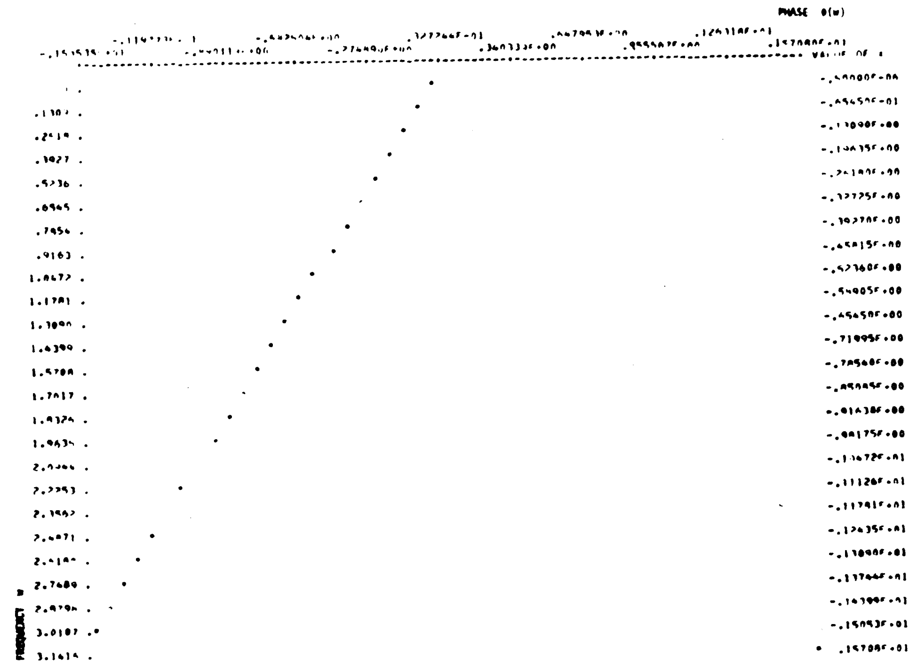
h(L) = (1 - L)²



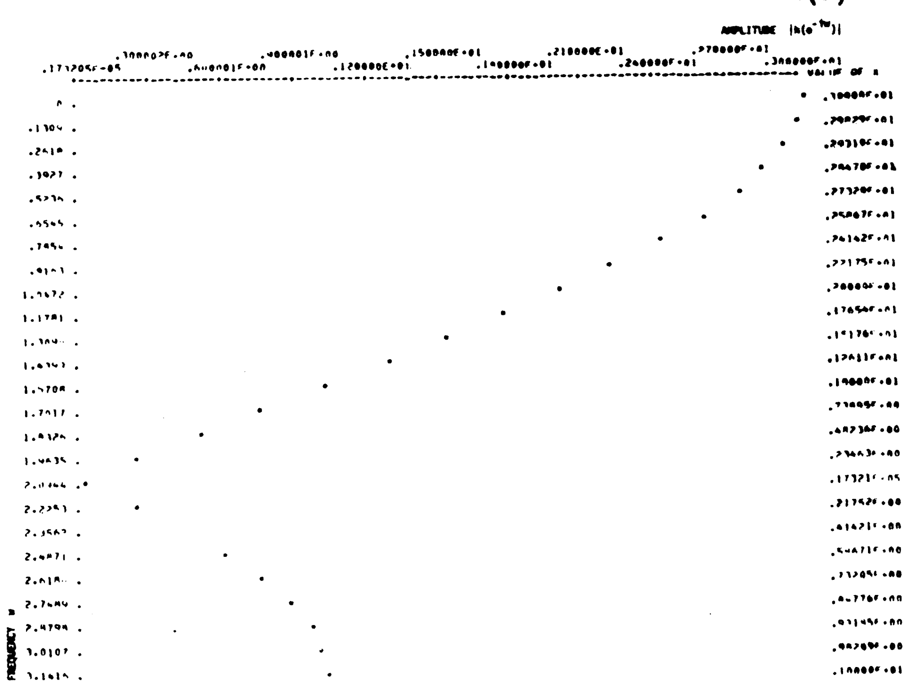
Page 2—Frequency Response Functions

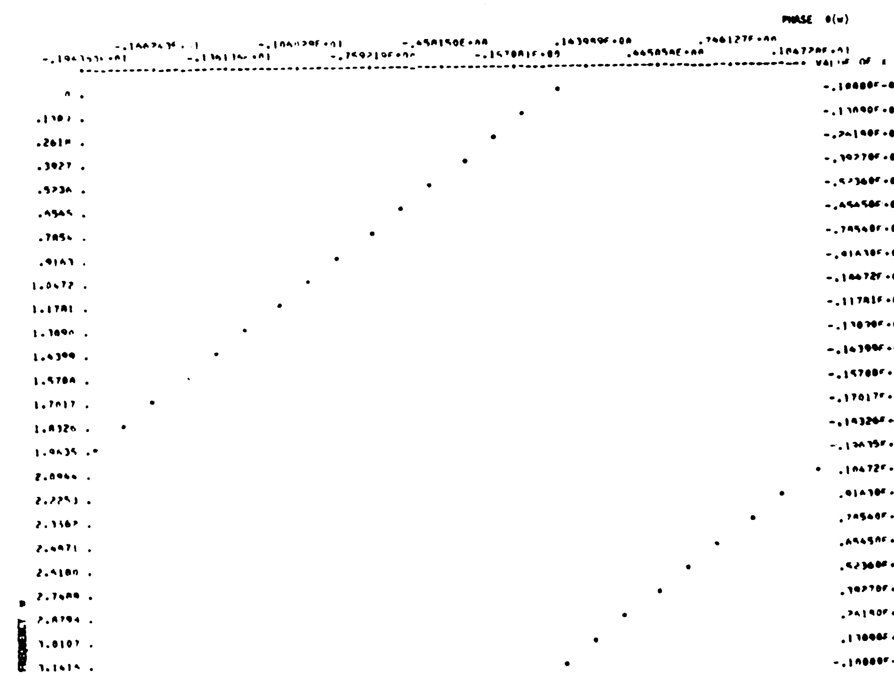
h(L) = 1 + L





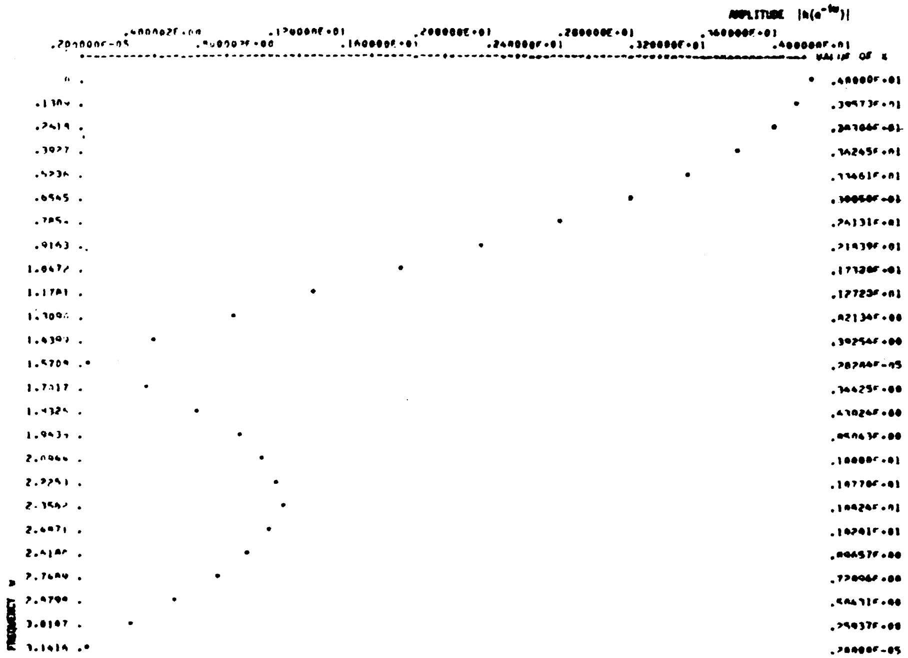
h(L) = 1 + L + L^2

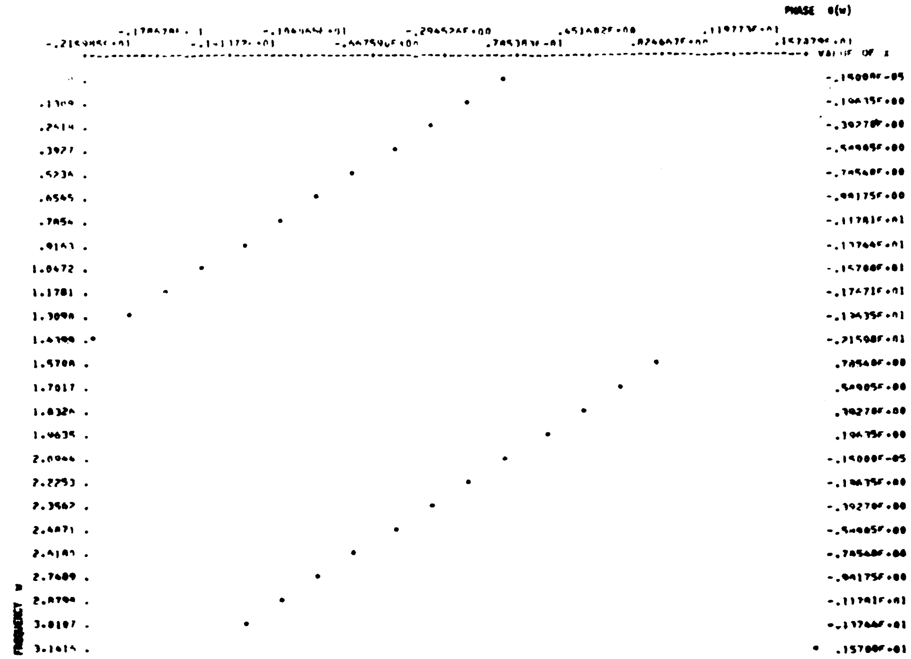




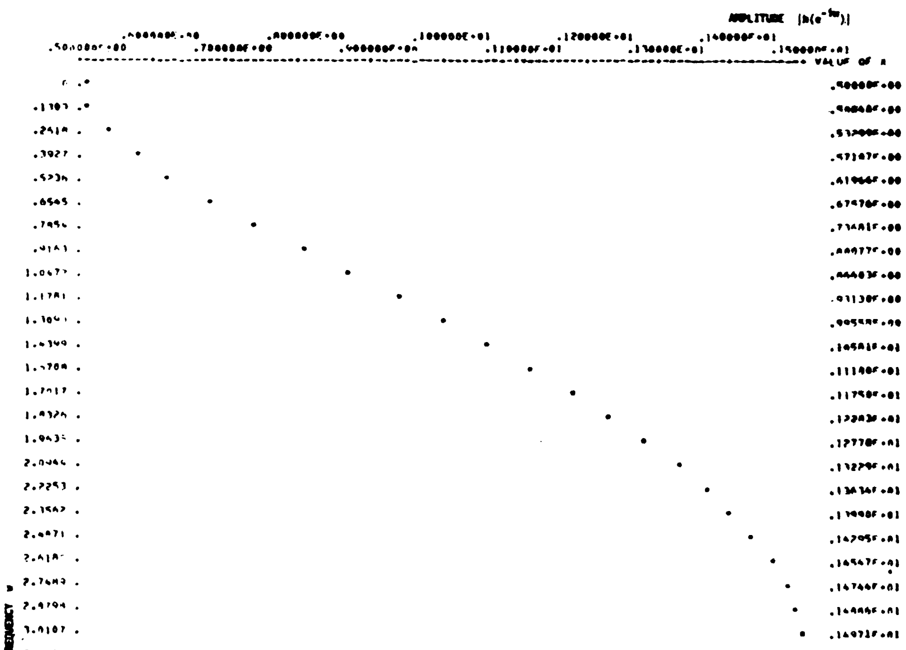
Page 3—Frequency Response Functions

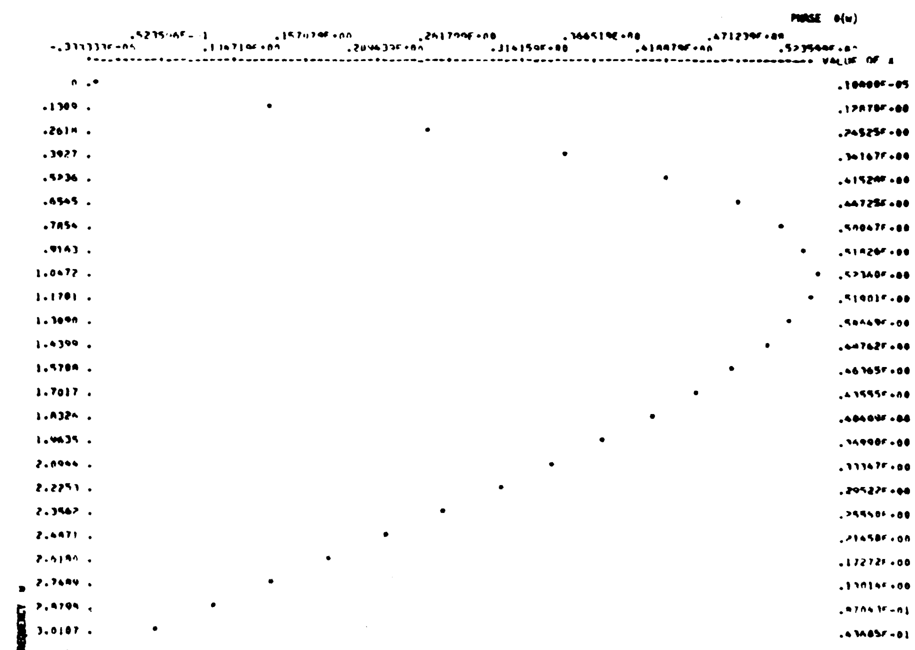
$h(L) = 1 + L + L^2 + L^3$





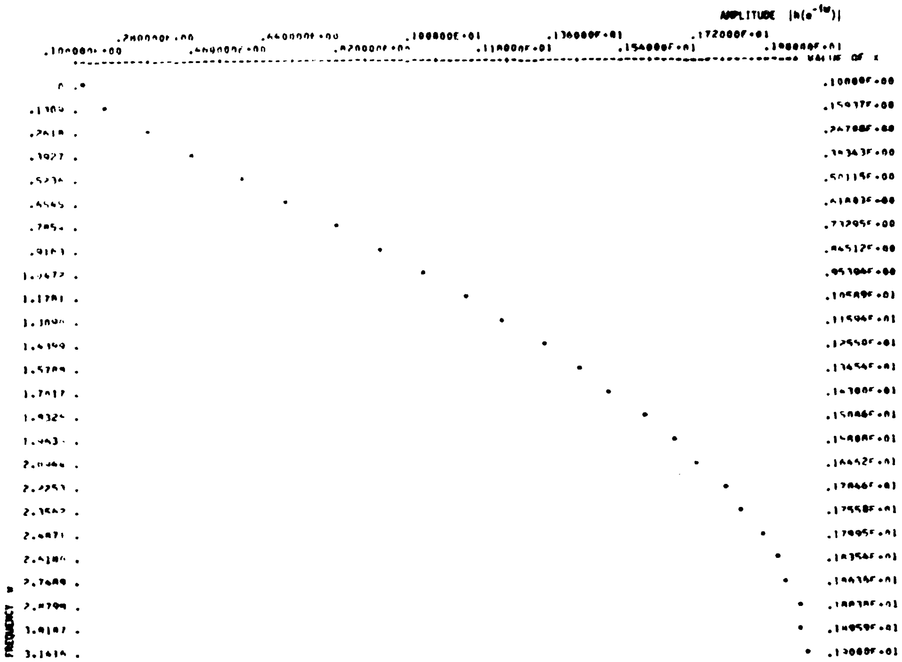
$h(L) = 1 - .5L$

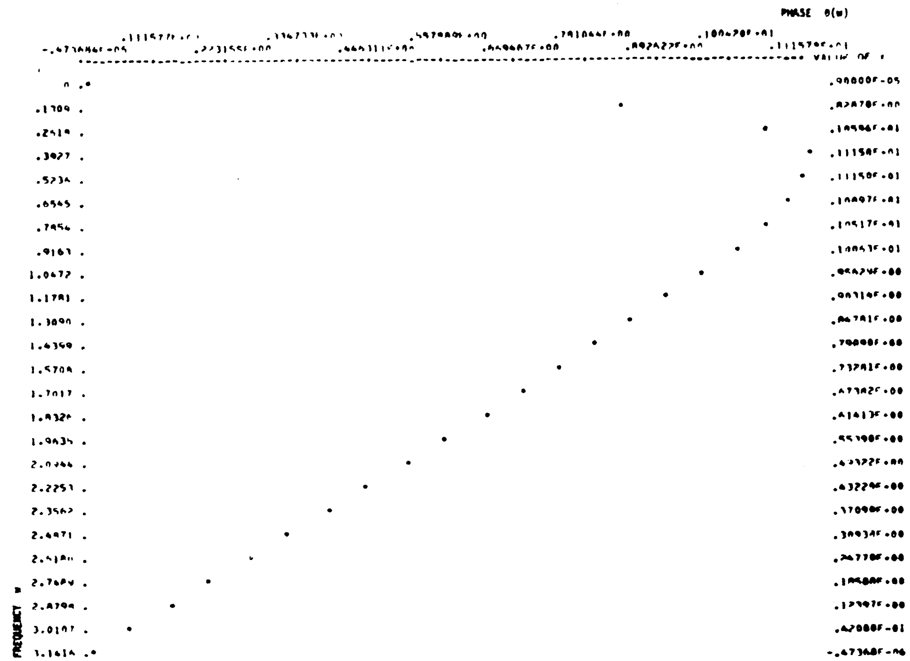




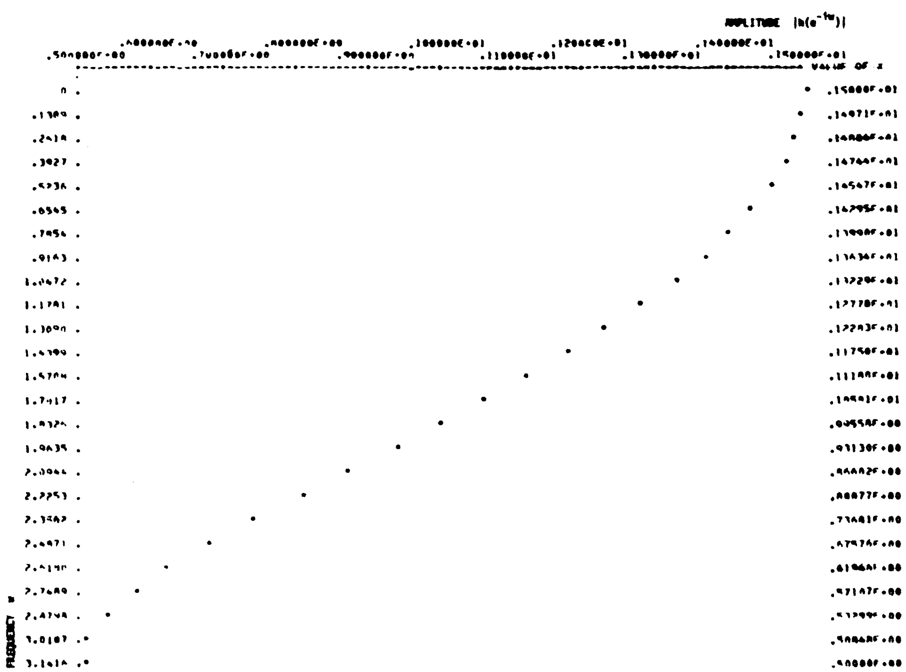
Page 4—Frequency Response Functions

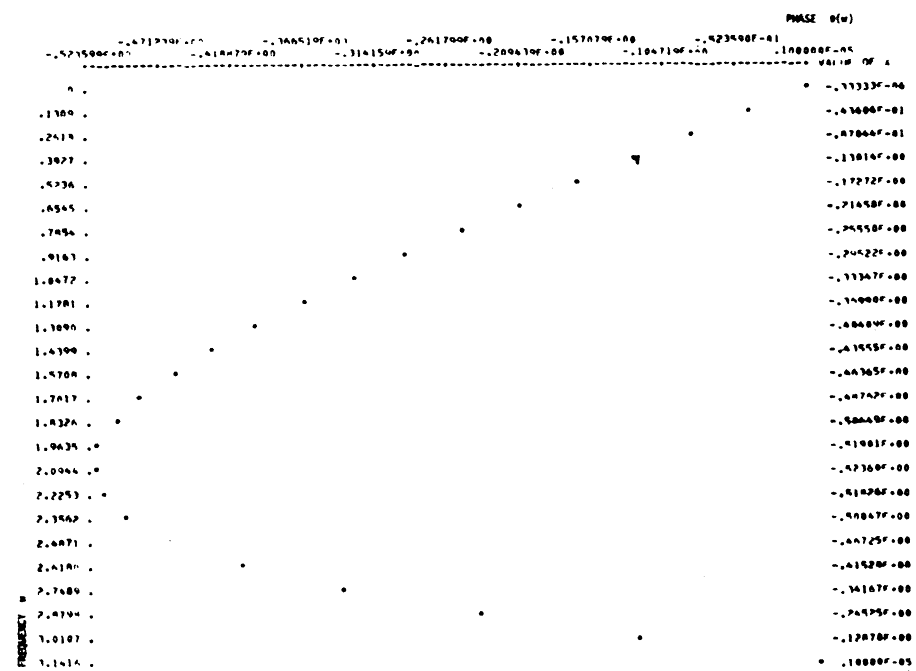
h(L) = 1 - .9L





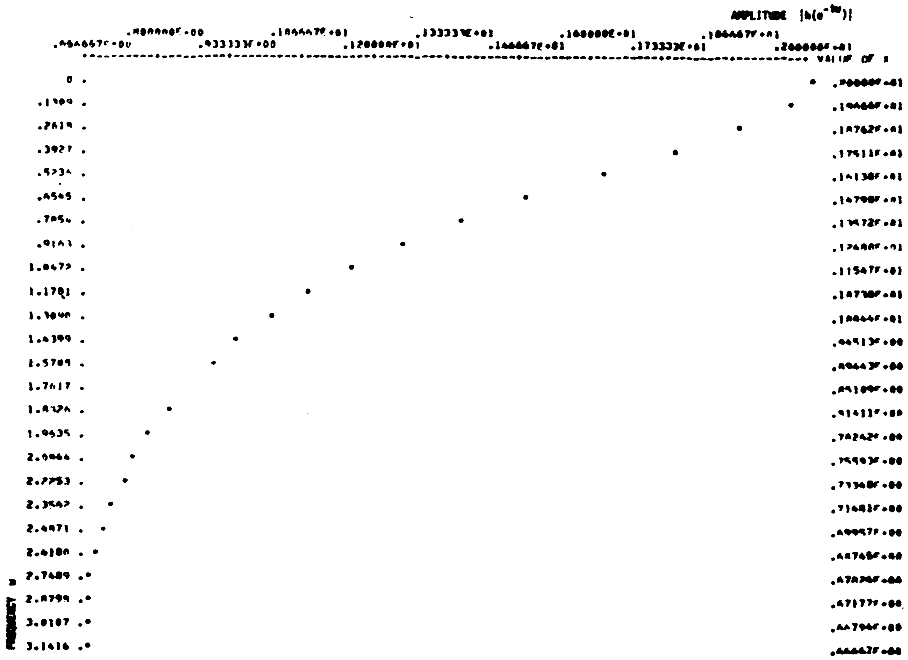
h(L) = 1 + .5L

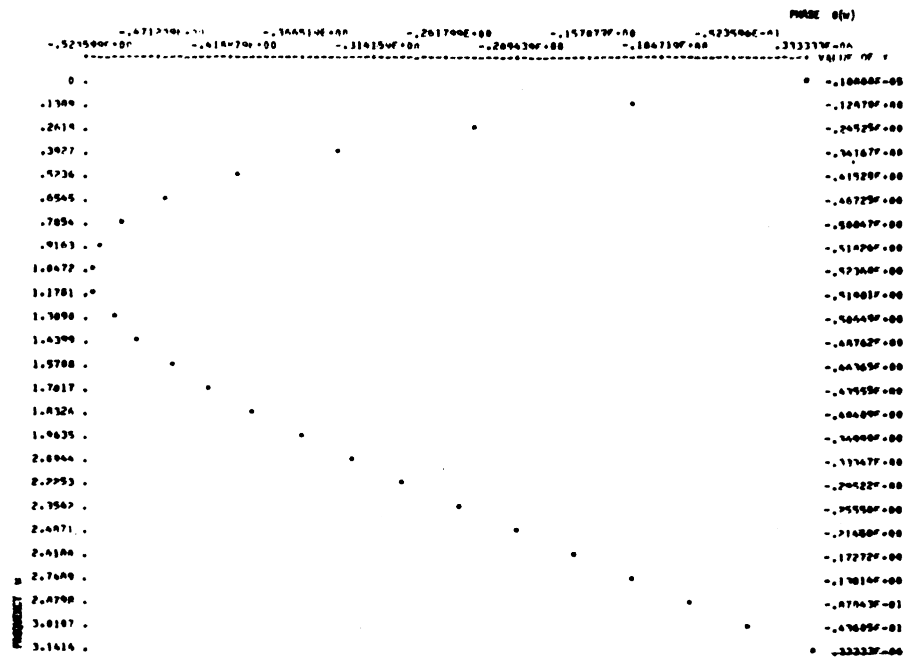




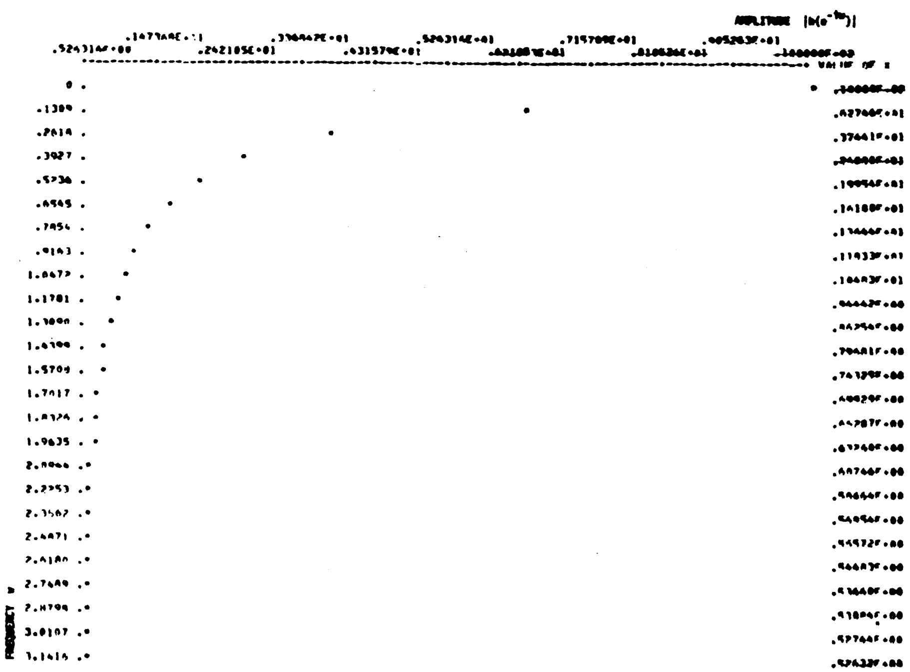
Page 5—Frequency Response Functions

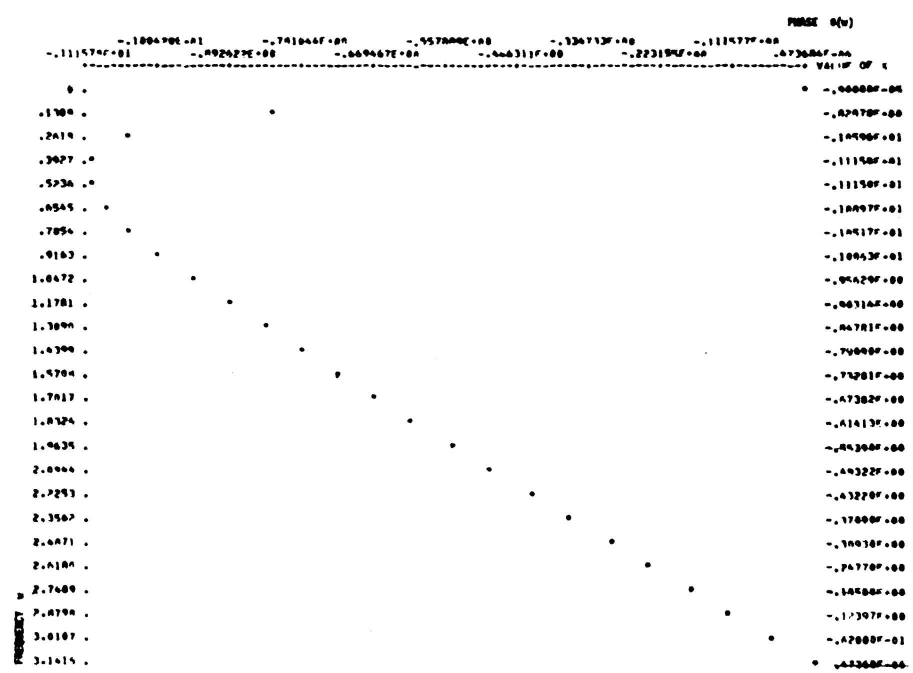
$$h(L) = (1 - .5L)^{-1}$$





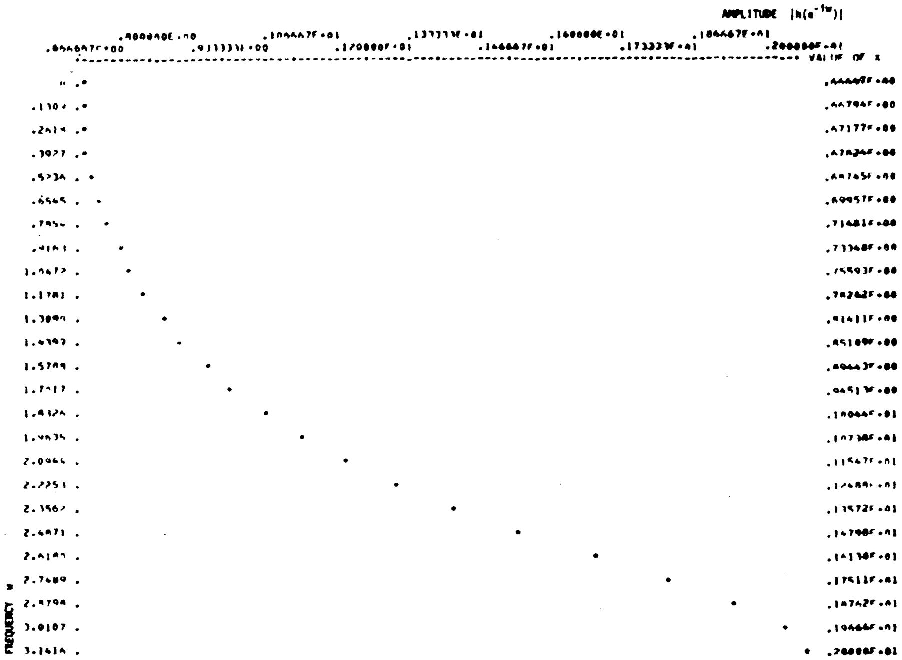
$$h(L) = (1 - .9L)^{-1}$$

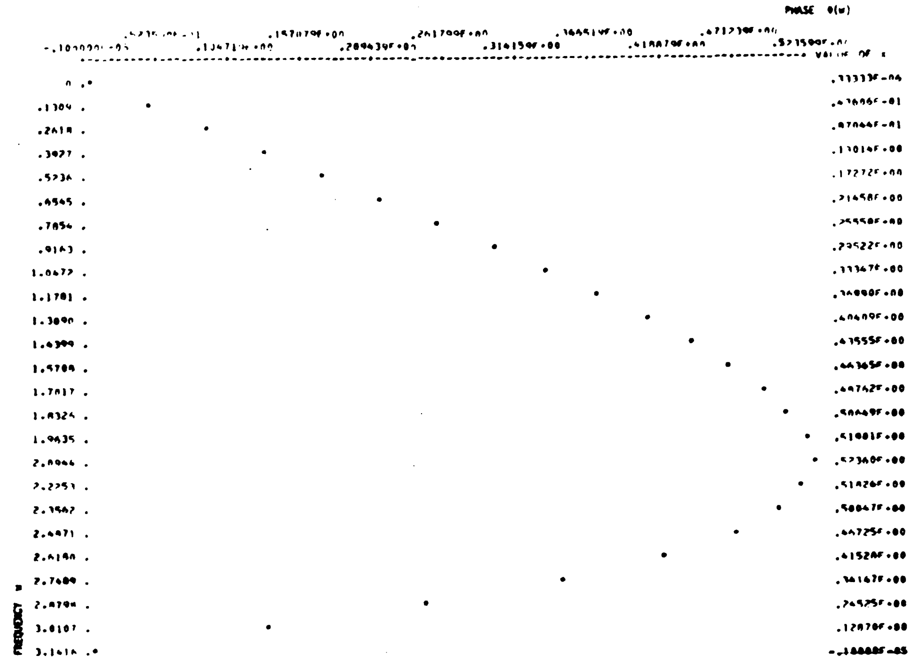




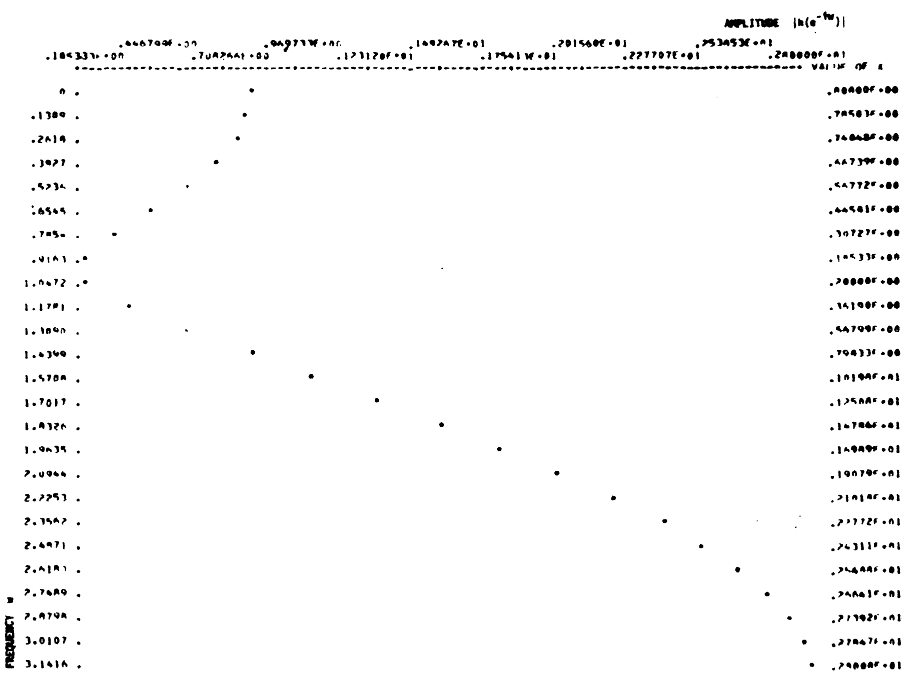
Page 6—Frequency Response Functions

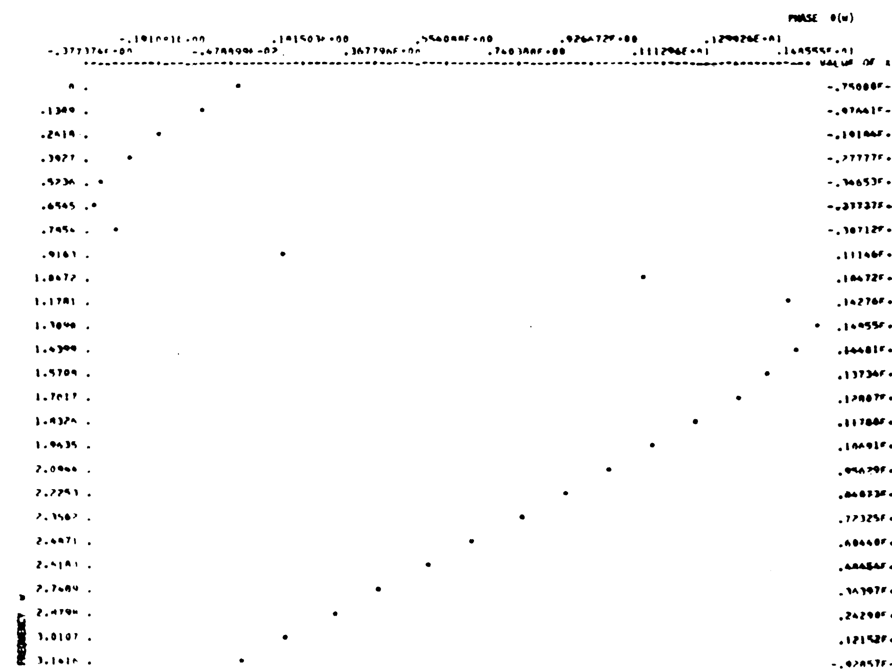
$$h(L) = (1 + .5L)^{-1}$$





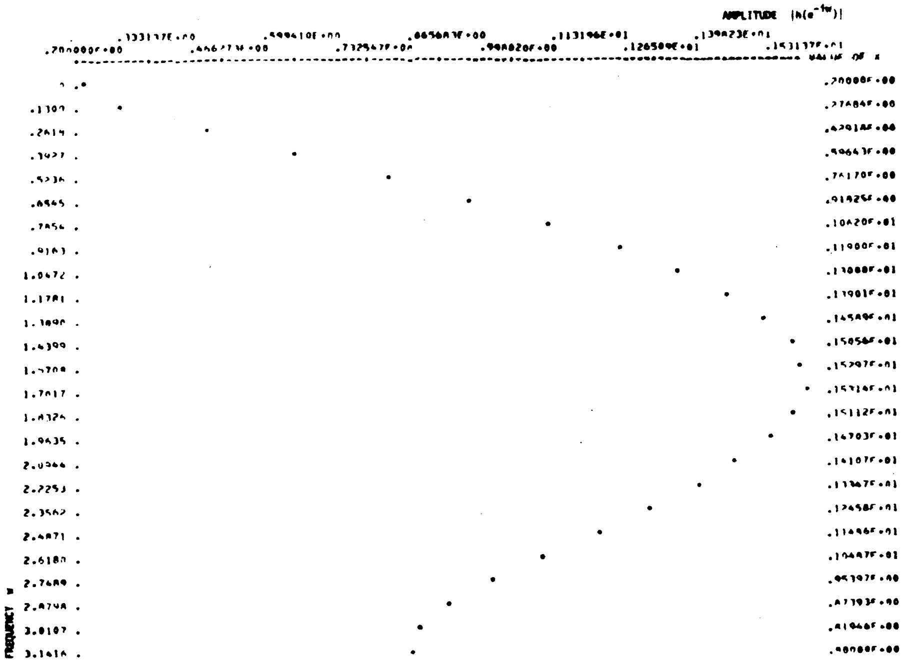
$$h(L) = 1 - L + .8L^2$$

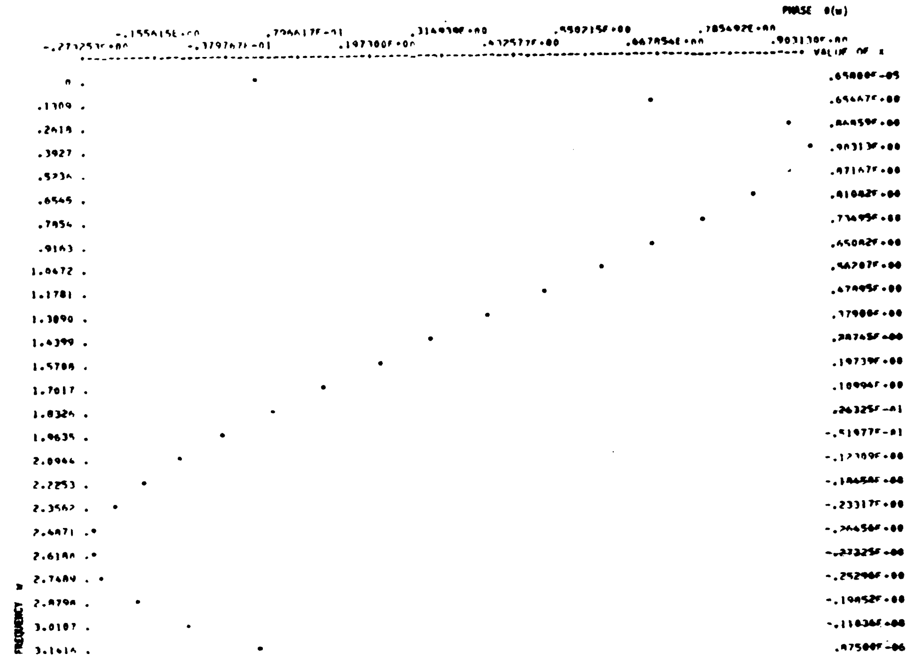




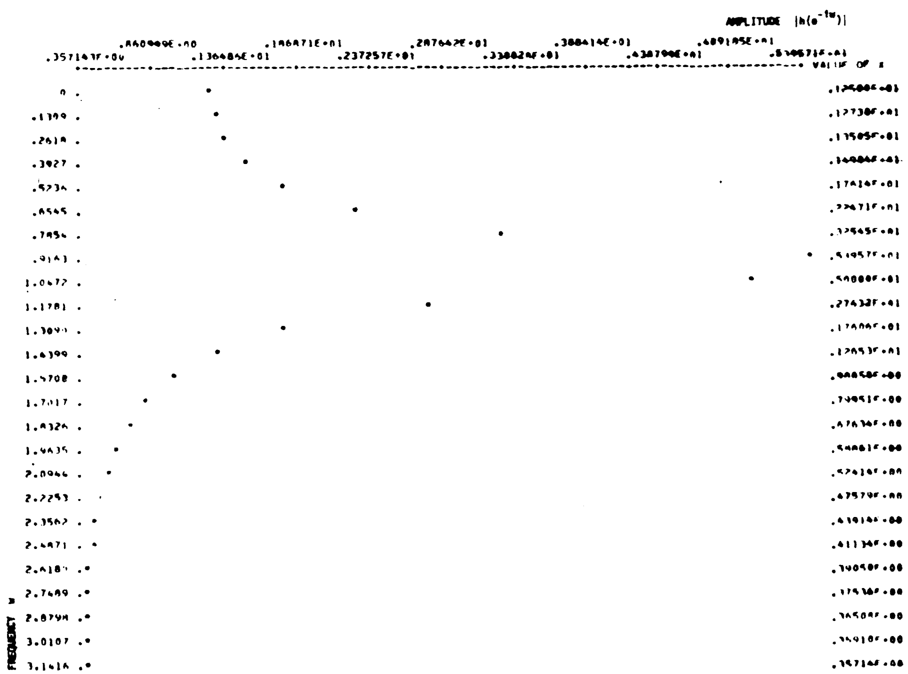
Page 7—Frequency Response Functions

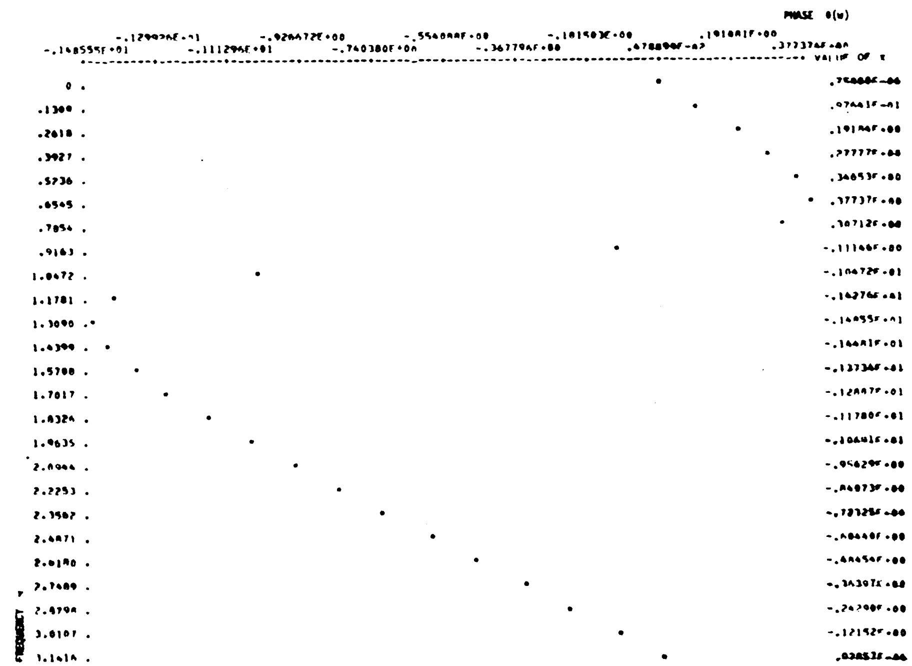
$h(L) = 1 - .3L - .5L^2$





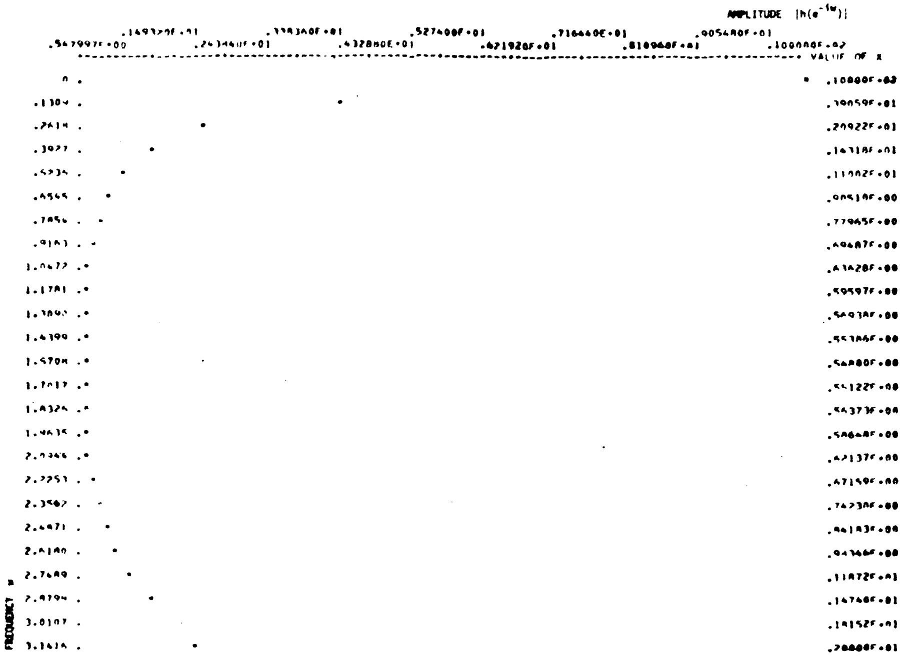
$h(L) = (1 - L + .8L^2)^{-1}$

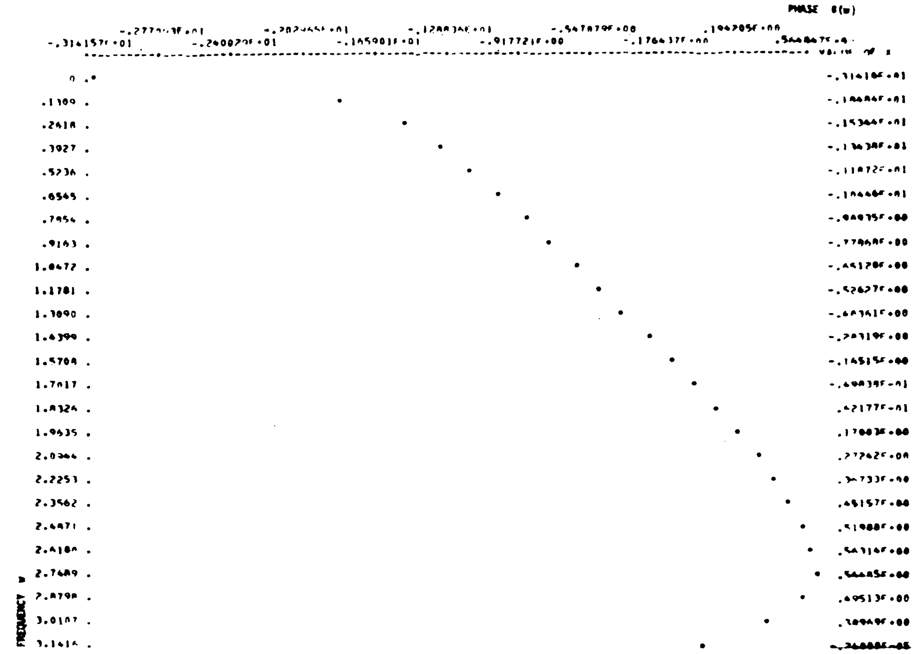




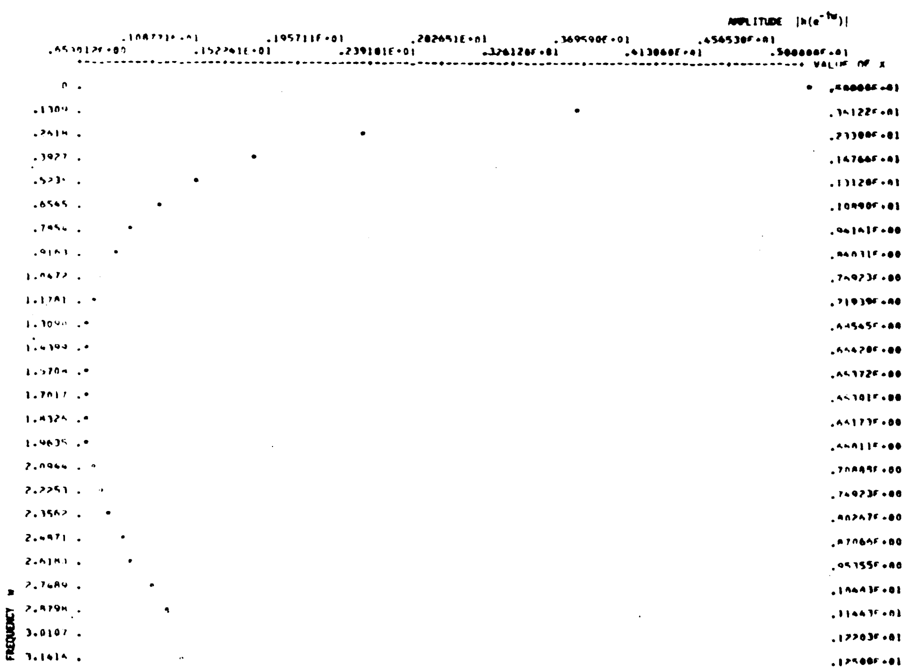
Page 8—Frequency Response Functions

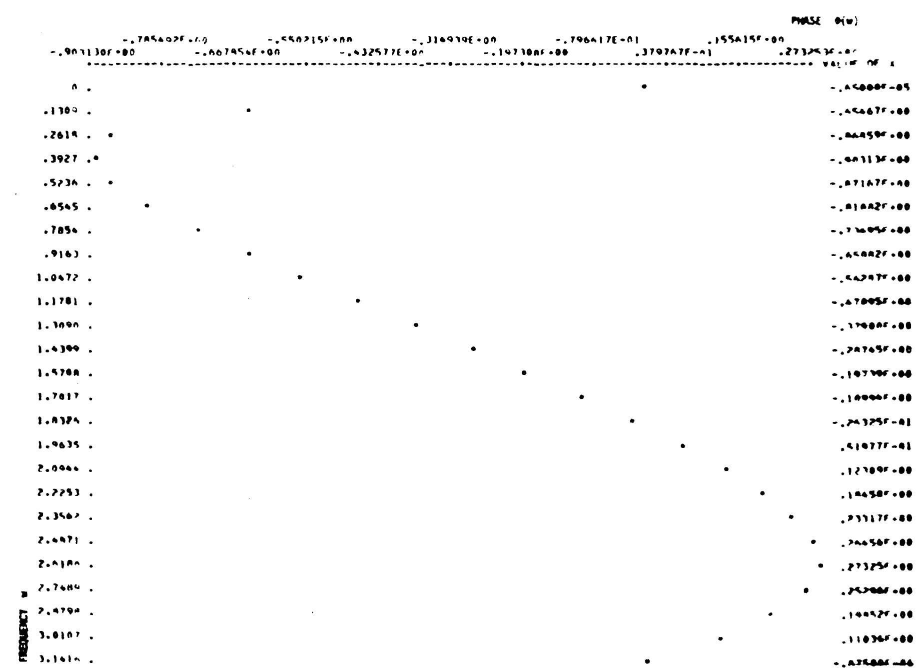
$h(L) = (1 - .3L - .8L^2)^{-1}$





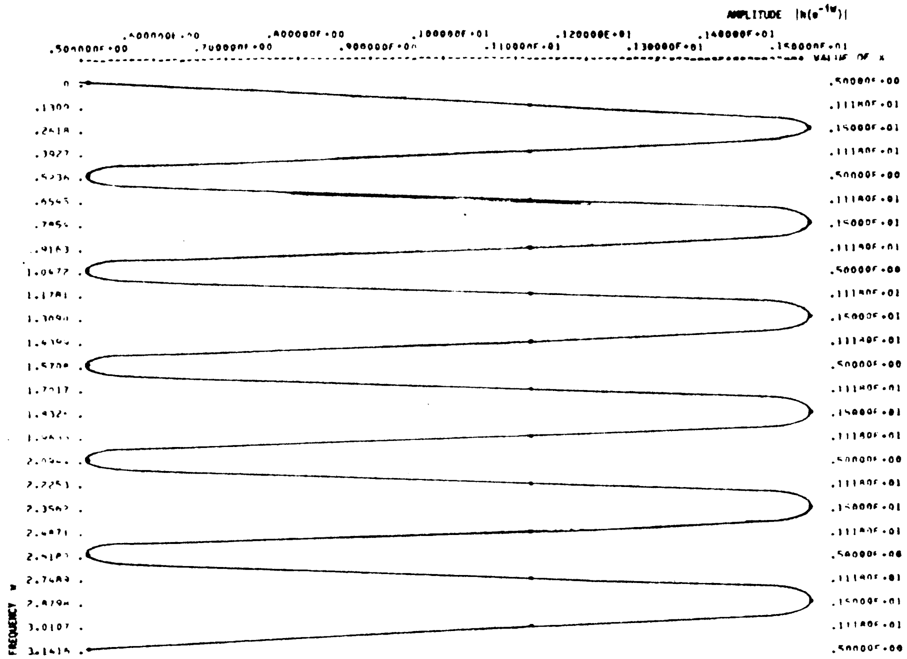
$h(L) = (1 - .3L - .5L^2)^{-1}$

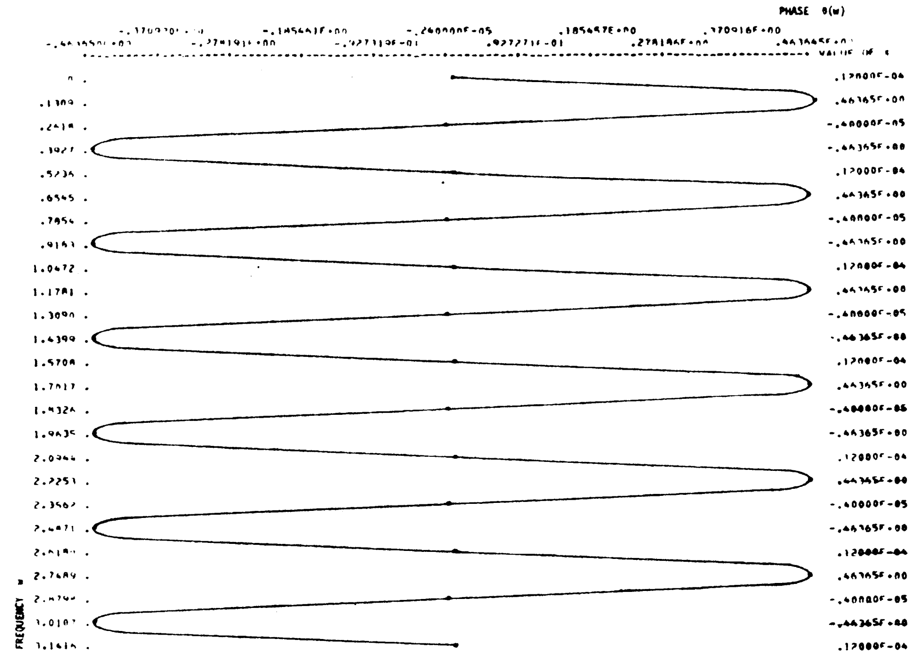




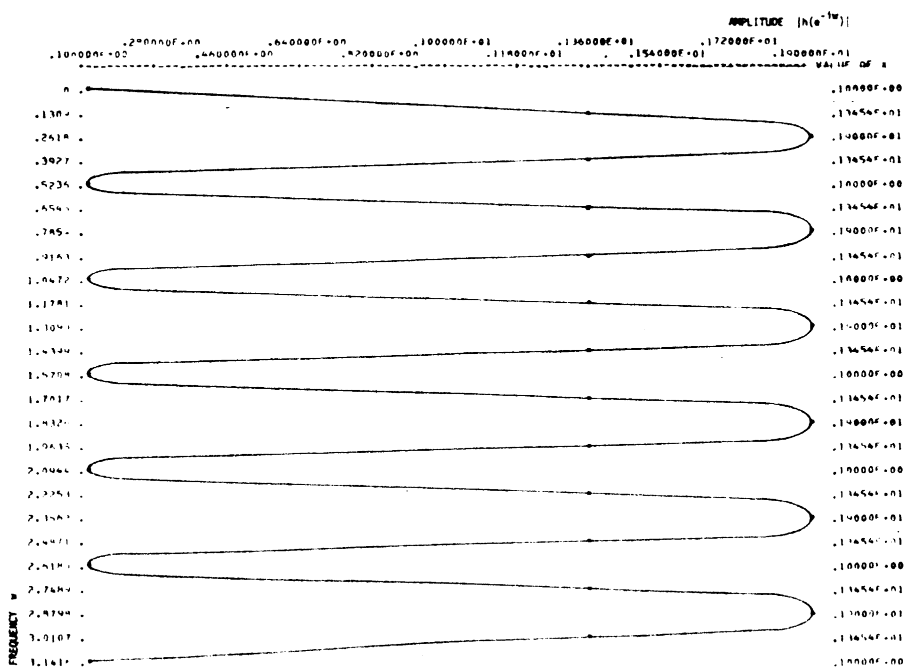
Page 9—Frequency Response Functions

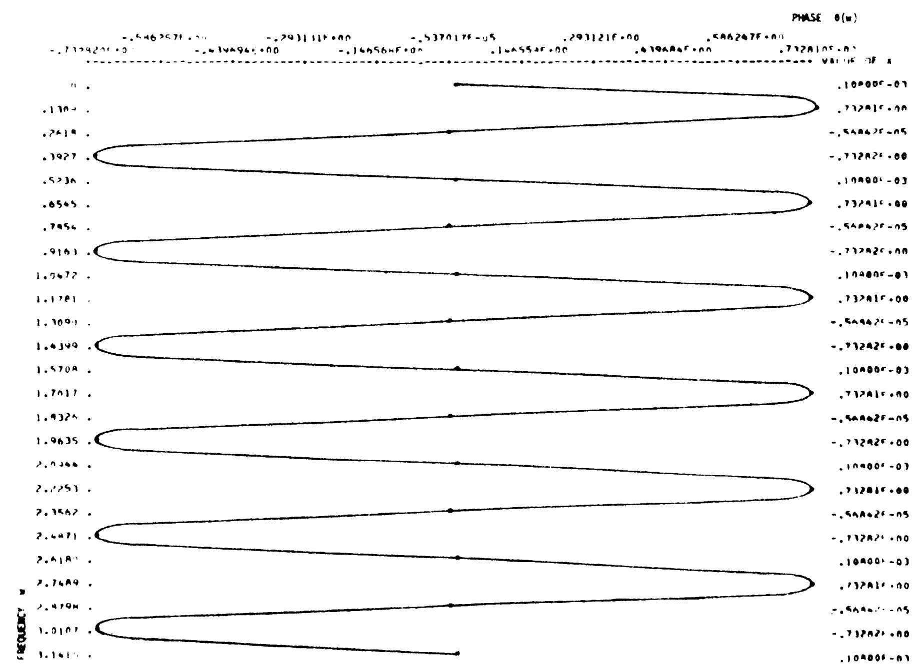
$h(L) = 1 - .5L^{12}$





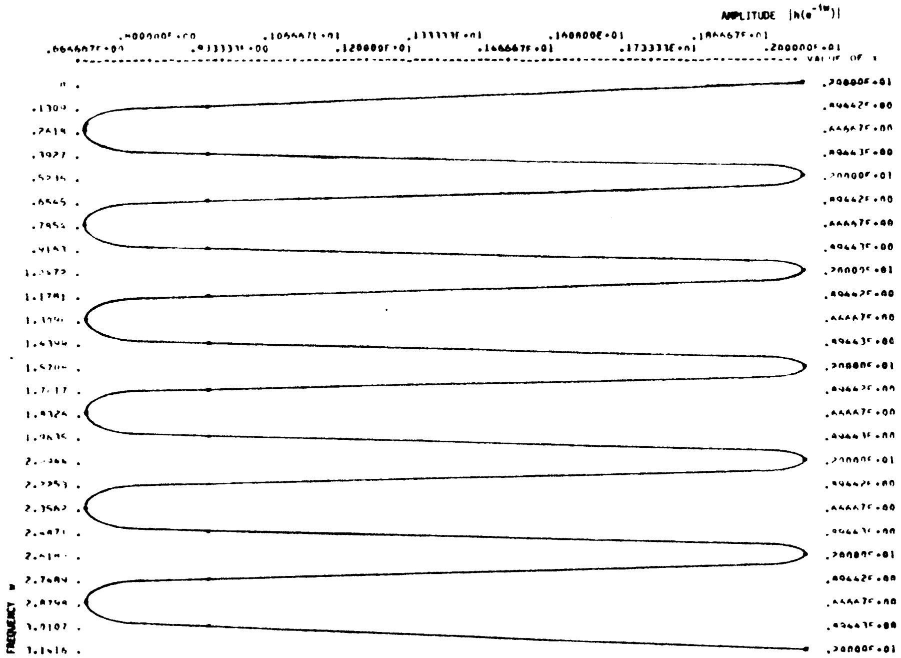
$h(L) = 1 - .9L^{12}$

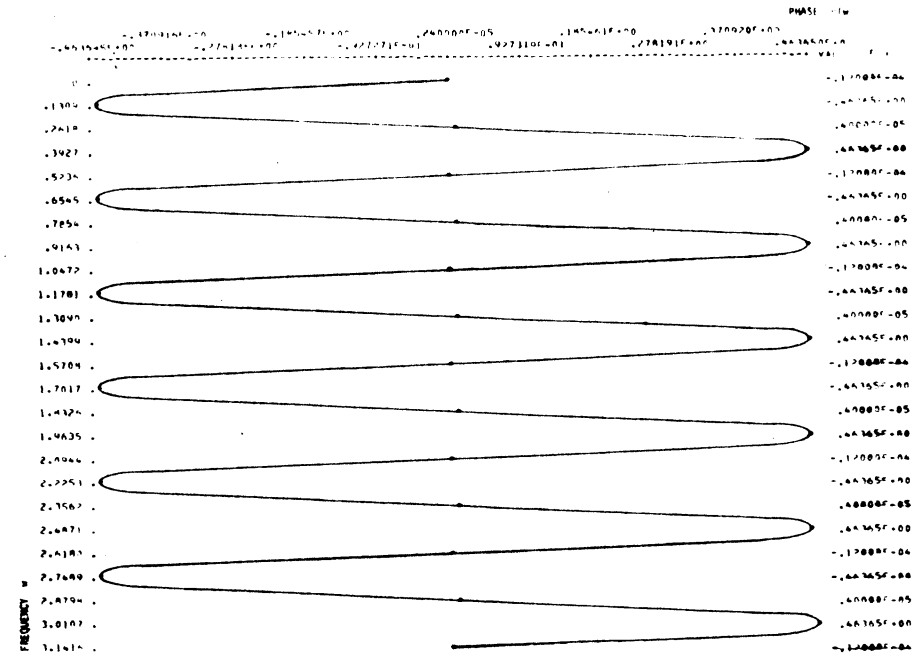




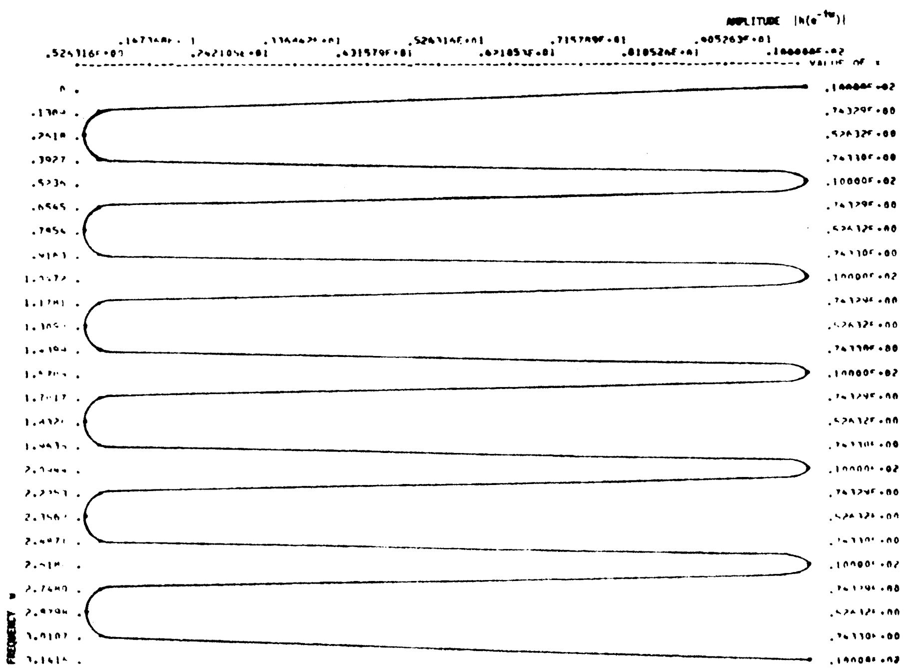
Page 10—Frequency Response Functions

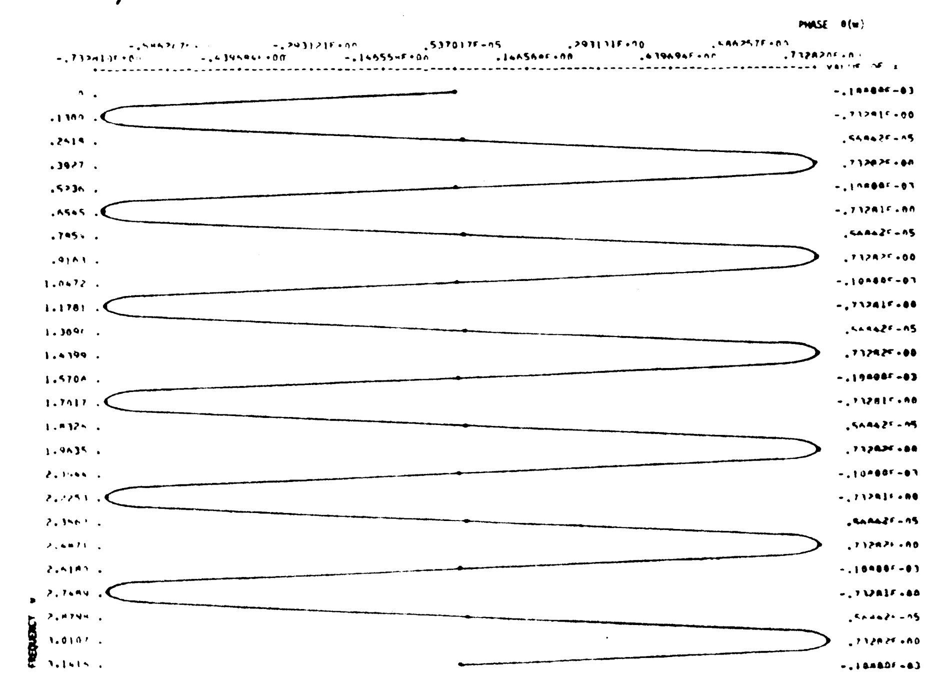
$h(L) = (1 - .5L^{12})^{-1}$





$h(L) = (1 - .9L^{12})^{-1}$





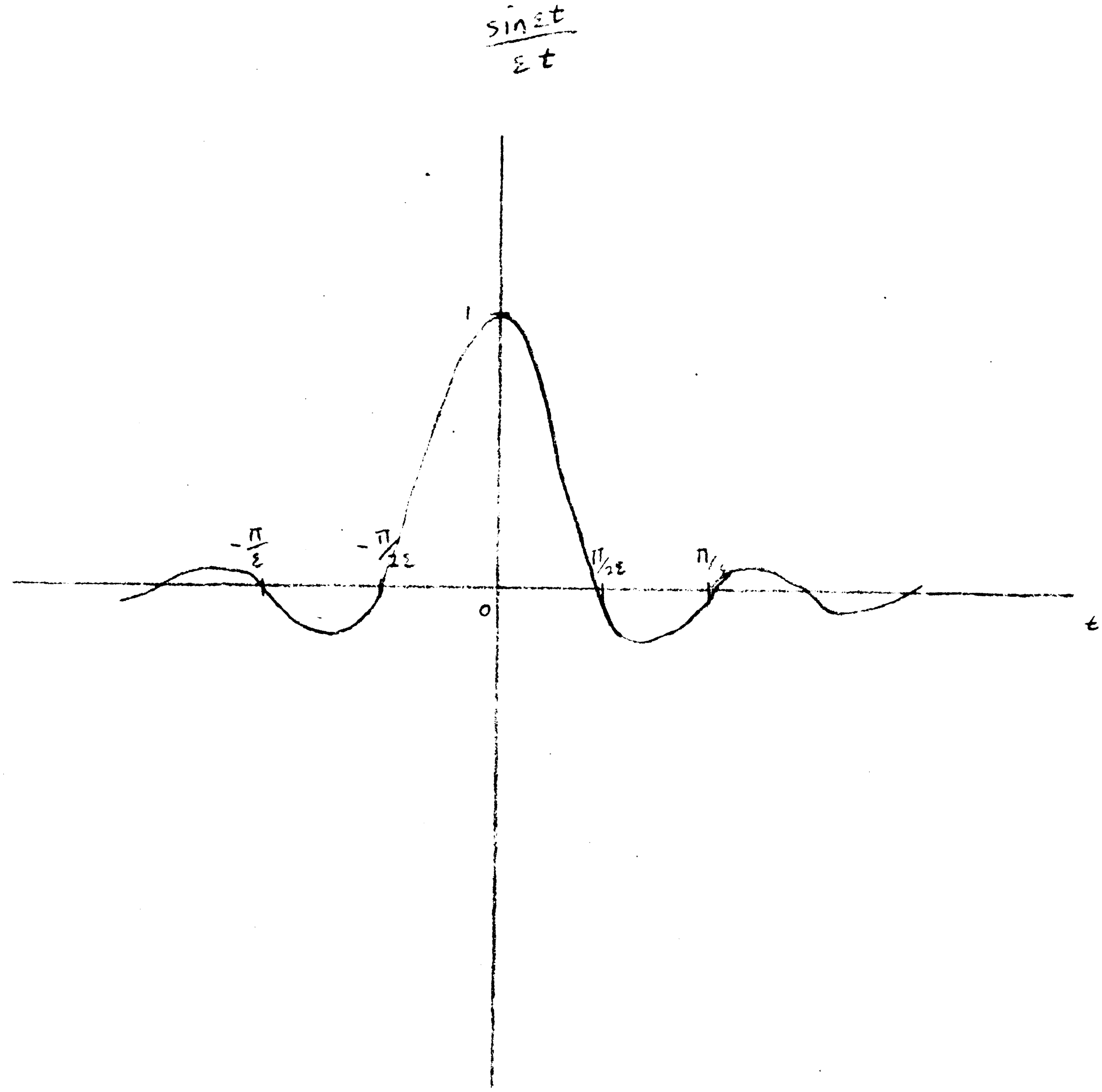
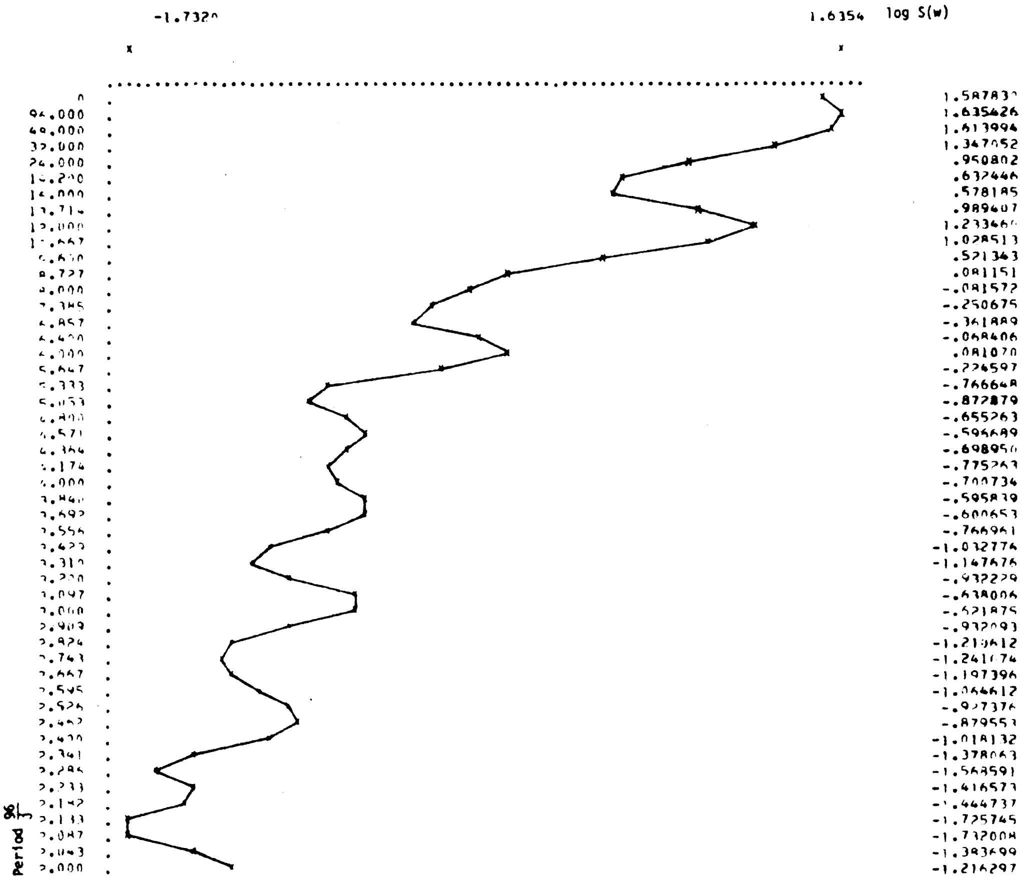


Figure 7

Plot of Estimated Spectrum of Call Rate



Plot of Estimated Spectrum of Time Rate

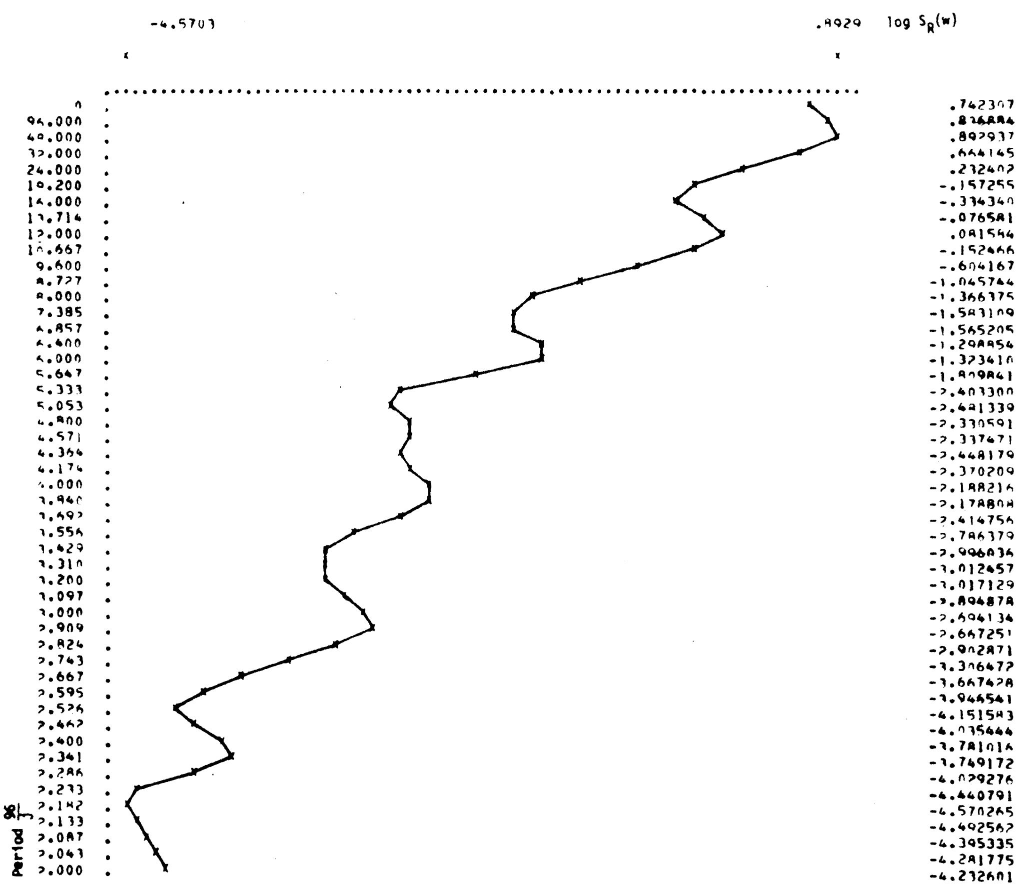


Figure 7

Table 1
Spectral Densities

Frequency* 1/38 C/UT	Power Spectra			Series			
	1	2	3	4	5	6	7
	y	$\bar{y}$	k	p	${}_tP^*_{t-1}$	r	m
0	.6725E-01	.6308E-01	.1009E+01	.1169E+01	.1169E+01	.5702E-02	.1600E+00
1	.3005E+00	.2001E-02	.3202E-01	.1775E+00	.1812E+00	.5208E-02	.1205E+00
2	.1595E+00	.2248E-03	.3597E-02	.2430E-01	.2627E-01	.2288E-02	.6955E-01
3	.9165E-01	.6507E-04	.1041E-02	.5964E-02	.7042E-02	.1264E-02	.4110E-01
4	.5787E-01	.2820E-04	.4511E-03	.2015E-02	.2652E-02	.7899E-03	.2640E-01
5	.3961E-01	.1535E-04	.2456E-03	.8460E-03	.1255E-02	.5397E-03	.1830E-01
6	.2892E-01	.9677E-05	.1548E-03	.4187E-03	.7033E-03	.3942E-03	.1348E-01
7	.2222E-01	.6771E-05	.1083E-03	.2369E-03	.4494E-03	.3033E-03	.1042E-01
8	.1781E-01	.5129E-05	.8206E-04	.1501E-03	.3201E-03	.2432E-03	.8373E-02
9	.1479E-01	.4145E-05	.6632E-04	.1053E-03	.2502E-03	.2019E-03	.6951E-02
10	.1265E-01	.3545E-05	.5671E-04	.8104E-04	.2126E-03	.1725E-03	.5930E-02
11	.1113E-01	.3197E-05	.5114E-04	.6811E-04	.1954E-03	.1513E-03	.5179E-02
12	.1005E-01	.3042E-05	.4866E-04	.6241E-04	.1936E-03	.1359E-03	.4619E-02
13	.9324E-02	.3068E-05	.4910E-04	.6245E-04	.2074E-03	.1250E-03	.4197E-02
14	.8906E-02	.3314E-05	.5303E-04	.6857E-04	.2411E-03	.1178E-03	.3878E-02
15	.8829E-02	.3893E-05	.6228E-04	.8337E-04	.3071E-03	.1142E-03	.3641E-02
16	.9224E-02	.5072E-05	.8115E-04	.1135E-03	.4336E-03	.1152E-03	.3470E-02
17	.1042E-01	.7431E-05	.1189E-03	.1737E-03	.6806E-03	.1230E-03	.3354E-02
18	.1278E-01	.1161E-04	.1858E-03	.2801E-03	.1115E-02	.1406E-03	.3287E-02
19	.1461E-01	.1476E-04	.2361E-03	.3602E-03	.1441E-02	.1546E-03	.3265E-02

*To get period, invert the integer in this column and then multiply by 38.

Table 2a
Coherence Between y and Various Series

Frequency* 1/38 C/UT	Coherence Series 1 and 2 $\bar{y}$	3 k	4 p	5 ${}_tP^*_{t-1}$	6 r	7 m
0	.9630E+00	.9630E+00	.8154E+00	.8154E+00	.9020E+00	0
1	.6551E+00	.6551E+00	.9892E+00	.9892E+00	.8012E+00	.5101E-01
2	.4309E+00	.4309E+00	.9837E+00	.9837E+00	.7152E+00	.4002E-01
3	.3741E+00	.3741E+00	.9727E+00	.9727E+00	.6898E+00	.3754E-01
4	.3590E+00	.3590E+00	.9576E+00	.9576E+00	.6793E+00	.3659E-01
5	.3527E+00	.3527E+00	.9390E+00	.9390E+00	.6738E+00	.3601E-01
6	.3489E+00	.3489E+00	.9176E+00	.9176E+00	.6706E+00	.3578E-01
7	.3472E+00	.3472E+00	.8944E+00	.8944E+00	.6687E+00	.3552E-01
8	.3489E+00	.3489E+00	.8702E+00	.8702E+00	.6678E+00	.3528E-01
9	.3552E+00	.3552E+00	.8461E+00	.8461E+00	.6676E+00	.3503E-01
10	.3671E+00	.3671E+00	.8229E+00	.8229E+00	.6682E+00	.3473E-01
11	.3858E+00	.3858E+00	.8017E+00	.8017E+00	.6695E+00	.3435E-01
12	.4123E+00	.4123E+00	.7836E+00	.7836E+00	.6719E+00	.3385E-01
13	.4480E+00	.4480E+00	.7696E+00	.7696E+00	.6758E+00	.3314E-01
14	.4938E+00	.4938E+00	.7616E+00	.7616E+00	.6820E+00	.3210E-01
15	.5508E+00	.5508E+00	.7618E+00	.7618E+00	.6922E+00	.3053E-01
16	.6192E+00	.6192E+00	.7738E+00	.7738E+00	.7091E+00	.2810E-01
17	.6967E+00	.6967E+00	.8009E+00	.8009E+00	.7369E+00	.2450E-01
18	.7706E+00	.7706E+00	.8382E+00	.8382E+00	.7744E+00	.2022E-01
19	.8048E+00	.8048E+00	.8586E+00	.8586E+00	.7962E+00	.1799E-01

(*See Table 1)

Table 2b
Coherence Between p and Various Series

Frequency* 1/38 C/UT	Series 5 $_tP^*_{t-1}$	6 r	7 m
0	.1000E+01	.8575E+00	.1368E+00
1	.1000E+01	.7810E+00	.4610E-01
2	.1000E+01	.6983E+00	.3880E-01
3	.1000E+01	.6686E+00	.3711E-01
4	.1000E+01	.6498E+00	.3627E-01
5	.1000E+01	.6336E+00	.3563E-01
6	.1000E+01	.6179E+00	.3505E-01
7	.1000E+01	.6025E+00	.3449E-01
8	.1000E+01	.5873E+00	.3393E-01
9	.1000E+01	.5729E+00	.3338E-01
10	.1000E+01	.5595E+00	.3285E-01
11	.1000E+01	.5479E+00	.3235E-01
12	.1000E+01	.5388E+00	.3188E-01
13	.1000E+01	.5332E+00	.3147E-01
14	.1000E+01	.5326E+00	.3110E-01
15	.1000E+01	.5398E+00	.3080E-01
16	.1000E+01	.5593E+00	.3055E-01
17	.1000E+01	.5974E+00	.3038E-01
18	.1000E+01	.6528E+00	.3027E-01
19	.1000E+01	.6857E+00	.3024E-01

(*See Table 1)

Table 3a
Gain of y on Various Series

Frequency* 1/38 C/UT	Series 1 on 2 $\bar{y}$	3 k	4 p	5 ${}_tP^*_{t-1}$	6 r	7 m
0	.1013E+01	.2533E+00	.2165E+00	.2165E+00	.3262E+01	0
1	.9917E+01	.2479E+01	.1294E+01	.1281E+01	.6799E+01	.3566E+00
2	.1749E+02	.4372E+01	.2541E+01	.2444E+01	.7062E+01	.3030E+00
3	.2295E+02	.5739E+01	.3866E+01	.3558E+01	.7073E+01	.2893E+00
4	.2714E+02	.6786E+01	.5244E+01	.4571E+01	.7054E+01	.2832E+00
5	.3017E+02	.7543E+01	.6631E+01	.5443E+01	.7032E+01	.2796E+00
6	.3229E+02	.8072E+01	.7960E+01	.6142E+01	.7013E+01	.2771E+00
7	.3376E+02	.8439E+01	.9160E+01	.6650E+01	.7000E+01	.2753E+00
8	.3481E+02	.8702E+01	.1016E+02	.6959E+01	.6993E+01	.2740E+00
9	.3559E+02	.8899E+01	.1090E+02	.7071E+01	.6993E+01	.2730E+00
10	.3620E+02	.9051E+01	.1134E+02	.6998E+01	.7001E+01	.2722E+00
11	.3666E+02	.9164E+01	.1145E+02	.6760E+01	.7019E+01	.2717E+00
12	.3692E+02	.9230E+01	.1124E+02	.6379E+01	.7050E+01	.2714E+00
13	.3690E+02	.9224E+01	.1072E+02	.5883E+01	.7100E+01	.2713E+00
14	.3643E+02	.9107E+01	.9945E+01	.5304E+01	.7182E+01	.2715E+00
15	.3535E+02	.8836E+01	.8982E+01	.4680E+01	.7315E+01	.2721E+00
16	.3356E+02	.8389E+01	.7929E+01	.4057E+01	.7536E+01	.2733E+00
17	.3125E+02	.7813E+01	.6931E+01	.3501E+01	.7899E+01	.2758E+00
18	.2912E+02	.7281E+01	.6184E+01	.3100E+01	.8390E+01	.2803E+00
19	.2822E+02	.7056E+01	.5901E+01	.2951E+01	.8674E+01	.2837E+00

(*See Table 1)

Table 3b
Gain of p on Various Series

Frequency* 1/38 C/UT	Series 4 on 1 y	2 $\bar{y}$	3 k	5 ${}_{t}P^{*}_{t-1}$	6 r	7 m
0	.3765E+01	.3961E+01	.9903E+00	.1000E+01	.1326E+02	.1000E+01
1	.7646E+00	.7481E+01	.1870E+01	.9899E+00	.5160E+01	.2606E+00
2	.3870E+00	.6820E+01	.1705E+01	.9617E+00	.2723E+01	.1164E+00
3	.2516E+00	.5949E+01	.1487E+01	.9203E+00	.1776E+01	.7338E-01
4	.1826E+00	.5215E+01	.1304E+01	.8716E+00	.1287E+01	.5261E-01
5	.1416E+00	.4599E+01	.1150E+01	.8209E+00	.9966E+00	.4059E-01
6	.1153E+00	.4106E+01	.1027E+01	.7716E+00	.8101E+00	.3300E-01
7	.9763E-01	.3735E+01	.9337E+00	.7260E+00	.6859E+00	.2800E-01
8	.8565E-01	.3473E+01	.8682E+00	.6849E+00	.6021E+00	.2467E-01
9	.7763E-01	.3309E+01	.8273E+00	.6487E+00	.5467E+00	.2249E-01
10	.7259E-01	.3233E+01	.8082E+00	.6173E+00	.5127E+00	.2119E-01
11	.7003E-01	.3236E+01	.8091E+00	.5905E+00	.4966E+00	.2062E-01
12	.6974E-01	.3315E+01	.8287E+00	.5677E+00	.4974E+00	.2076E-01
13	.7179E-01	.3465E+01	.8662E+00	.5488E-00	.5162E+00	.2164E-01
14	.7657E-01	.3679E+01	.9198E+00	.5333E+00	.5569E+00	.2345E-01
15	.8482E-01	.3945E+01	.9863E+00	.5210E+00	.6277E+00	.2655E-01
16	.9759E-01	.4236E+01	.1059E+01	.5117E+00	.7425E+00	.3162E-01
17	.1155E+00	.4509E+01	.1127E+01	.5052E+00	.9183E+00	.3966E-01
18	.1355E+00	.4708E+01	.1177E+01	.5013E+00	.1140E+01	.5079E-01
19	.1455E+00	.4781E+01	.1195E+01	.5000E+00	.1264E+01	.5775E-01

(*See Table 1)

Table 3c
Gain of r on Various Series

Frequency* 1/38 C/UT	Series 6 on 1 y	2 $\bar{y}$	3 k	4 p	5 $t^{P^*}_{t-1}$	7 m
0	.2765E+00	.2948E+00	.7370E-01	.6466E-01	.6466E-01	0
1	.1178E+00	.1549E+01	.3871E+00	.1514E+00	.1498E+00	.6910E-01
2	.1013E+00	.2885E+01	.7211E+00	.2564E+00	.2466E+00	.7087E-01
3	.9753E-01	.3751E+01	.9377E+00	.3764E+00	.3464E+00	.7124E-01
4	.9629E-01	.4239E+01	.1060E+01	.5047E+00	.4399E+00	.7148E-01
5	.9581E-01	.4478E+01	.1119E+01	.6358E+00	.5219E+00	.7170E-01
6	.9562E-01	.4559E+01	.1140E+01	.7628E+00	.5886E+00	.7193E-01
7	.9554E-01	.4543E+01	.1136E+01	.8784E+00	.6377E+00	.7219E-01
8	.9550E-01	.4466E+01	.1117E+01	.9754E+00	.6681E+00	.7249E-01
9	.9548E-01	.4352E+01	.1088E+01	.1048E+01	.6798E+00	.7288E-01
10	.9544E-01	.4215E+01	.1054E+01	.1091E+01	.6738E+00	.7337E-01
11	.9539E-01	.4060E+01	.1015E+01	.1103E+01	.6515E+00	.7404E-01
12	.9531E-01	.3888E+01	.9721E+00	.1083E+01	.6150E+00	.7497E-01
13	.9517E-01	.3699E+01	.9247E+00	.1033E+01	.5669E+00	.7632E-01
14	.9496E-01	.3487E+01	.8717E+00	.9564E+00	.5101E+00	.7839E-01
15	.9463E-01	.3249E+01	.8122E+00	.8599E+00	.4480E+00	.8172E-01
16	.9409E-01	.2991E+01	.7477E+00	.7532E+00	.3854E+00	.8740E-01
17	.9328E-01	.2735E+01	.6838E+00	.6505E+00	.3286E+00	.9729E-01
18	.9231E-01	.2535E+01	.6337E+00	.5724E+00	.2870E+00	.1125E+00
19	.9179E-01	.2457E+01	.6141E+00	.5425E+00	.2712E+00	.1226E+00

(*See Table 1)

References

Anderson, T.W., <u>The Statistical Analysis of Time Series</u>, Wiley, 1970.

Chow, Gregory C., and R.E. Levitan, "Nature of Business Cycles Implicit in a Linear Economic Model," <u>Quarterly Journal of Economics</u>, August 1969.

Granger, C.W.J., "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," <u>Econometrica</u>, 37 (1969), 424-438.

Howrey, E.P., "A Spectrum Analysis of the Long-Swing Hypothesis," <u>International Economic Review</u>, June 1968.

Howrey, E.P., "Stochastic Properties of the Klein-Goldberger Model," <u>Econometrica</u>, January 1971.

Koopmans, L.H., <u>The Spectral Analysis of Time Series</u>, Academic Press, 19

Papoulis, A., <u>Probability, Random Variables, and Stochastic Processes</u>, McGraw-Hill, 19

Sims, C.A., "Money, Income, and Causality," <u>American Economic Review</u>, September 1972.

Whittle, P., <u>Prediction and Regulation by Linear Least Square Methods</u>, Van Nostrand, 1963.