

## CHAPTER XI

# LINEAR STOCHASTIC DIFFERENCE EQUATIONS

### 1. INTRODUCTION

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 do not 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 displays data on six time series for the postwar U.S.: real GNP, the unemployment rate, the Baa bond rate, the percentage rate of change in the real money supply, the inflation rate in the GNP deflator, real output (GNP)

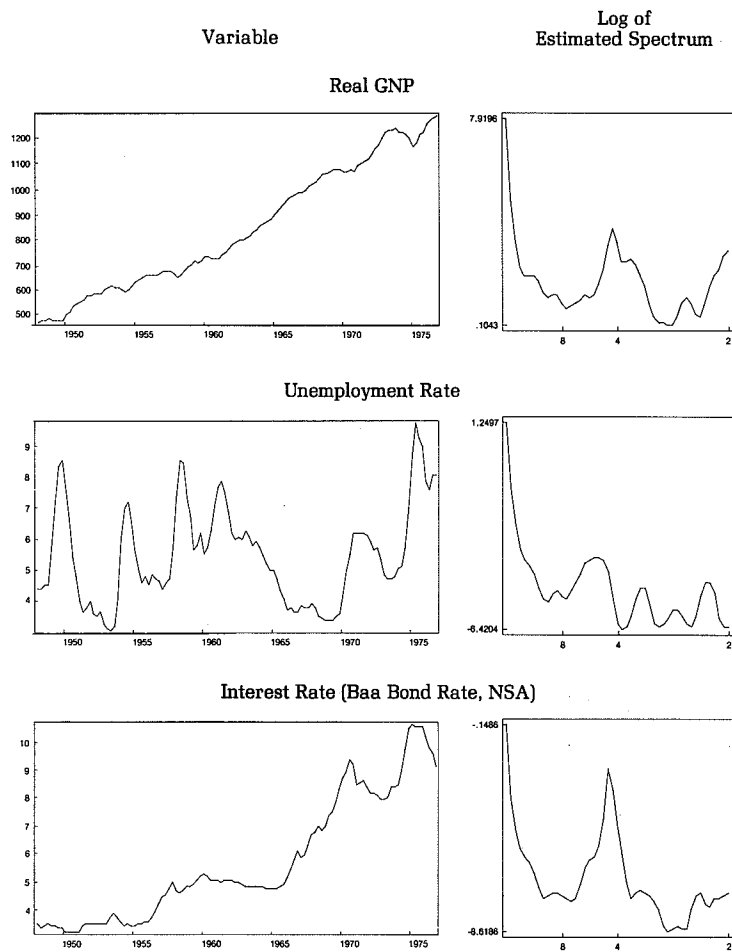
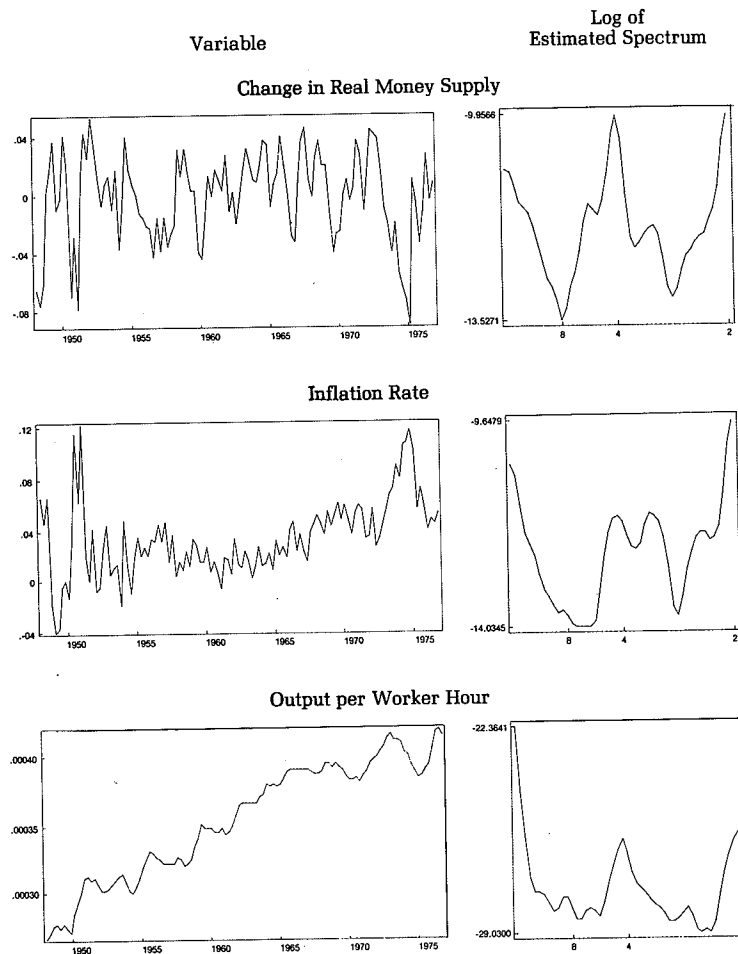
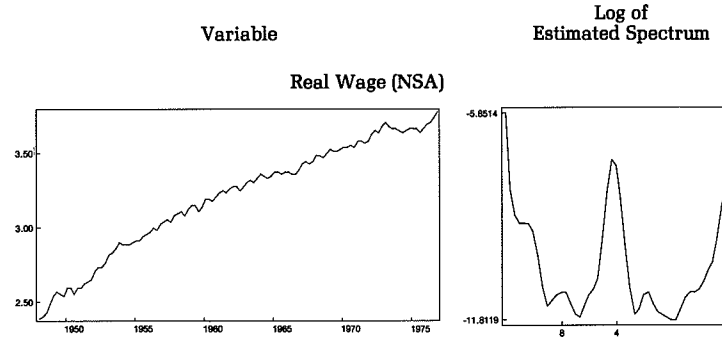


FIGURE 1





per man-hour, and an estimate of the straight time real wage.<sup>1</sup> The graphs labeled “log of estimated spectrum” are included for later reference. The “business cycle” shows up in the irregular, cumulative movements in real GNP, unemployment, and output per man-hour. There are recessions in 1954, 1958, 1960, 1970, and 1974–1975, these being characterized by reductions in real GNP and substantial increases in the unemployment rate. Notice that output per man-hour is markedly “procyclical,” i.e., its variation over the business cycle roughly matches that of GNP. This pattern would not be predicted by a straightforward application of the law of diminishing returns since the employment capital ratio itself is procyclical. In Chapter XVI we shall study one possible explanation of the paradoxical cyclical behavior of output per man-hour. Notice that the “cycles” in these variables are irregular in length and do not “look like” those generated by low-order nonstochastic difference equations.

As we have seen, low-order nonstochastic linear 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

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

<sup>1</sup> The data are from the NBER’s Troll data bank. The data are for the civilian nonagricultural unemployment rate; real GNP; Moody’s Baa rate, not seasonally adjusted; inflation rate in the GNP price deflator; the real money supply measured as currency plus demand deposits (M1) divided by the GNP price deflator; output per man-hour measured by real GNP divided by (employment on nonagricultural payrolls, private and government, multiplied by average weekly hours in manufacturing); the real wage measured as average straight time hourly earnings in manufacturing, not seasonally adjusted, divided by the GNP deflator. The spectral densities are for the period 1948I–1976IV.

where the  $\lambda_j$  are the roots of the characteristic equation and the  $a_j$  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 parametrization. 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 linear difference equations do not 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 linear 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 macroeconomic models, an idea that was introduced into economics by Slutsky (1937) and Frisch (1933). In addition to underlying the Slutsky–Frisch framework for business cycle analysis, linear stochastic difference equations provide the foundations for two important recent developments in macroeconomics: Sims’s method of studying the exogeneity or “causal” structure among sets of time series, and the construction of stochastic rational expectations models of the kind pioneered by Muth and Lucas. This chapter describes the elements of linear stochastic difference equations and some of their applications in economics.<sup>2</sup>

## 2. PRELIMINARY CONCEPTS

A *stochastic process* is a collection of random variables, a collection indexed by a variable  $t$ . In our work we shall regard  $t$  as time and will require  $t$  to be an integer, so that we shall be working in discrete time. Thus, the stochastic process  $y_t$  is a collection of random variables  $\dots, y_{-1}, y_0, y_1, y_2, \dots$ , 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 shall 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

<sup>2</sup> The reader is assumed to be familiar with complex variables. The chapter on complex variables in Allen (1960) is a good reference.

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  $y$  is not normal, the first and second moments contain much 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

$$E y_t = \mu_t, \quad t \in T,$$

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 *stationary* (or covariance stationary or second-order stationary) if  $\mu_t$  is independent of  $t$  and if  $\sigma_{t,s}$  depends only on  $t - s$ . We shall henceforth deal with such stationary processes. The first and second moments of a stationary process are summarized by the mean  $\mu$  and the *covariogram*  $c(\tau)$  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  $\tau = t - s$ . The covariogram is easily verified to be symmetric, i.e.,  $c(\tau) = c(-\tau)$ , and to obey  $c(0) \geq |c(\tau)|$  for all  $\tau$ , 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  $a_j$  are fixed real numbers and  $t_1, \dots, t_n$  are integers. We must require that the random variable  $y$  have nonnegative variance, so that

$$\begin{aligned} E y^2 &= E \left( \sum_{j=1}^n a_j x_{t_j} \sum_{k=1}^n a_k x_{t_k} \right) = \sum_{j=1}^n \sum_{k=1}^n a_j a_k E x_{t_k} x_{t_j} \\ &= \sum_{j=1}^n \sum_{k=1}^n a_j a_k c(t_k - t_j) \geq 0. \end{aligned}$$

This last inequality is required to hold for any  $n$ , any list of  $a_j$ , and any selection of  $(t_1, t_2, \dots, t_n)$ . A sequence  $c(\tau)$  that satisfies this condition is said to be "non-negative 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.<sup>3</sup>

<sup>3</sup> 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.

A basic building block is the serially uncorrelated random process  $\varepsilon_t$ , which satisfies:

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

This process is (wide-sense) stationary, each variate being uncorrelated with itself lagged  $s = \pm 1, \pm 2, \dots$  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.

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

$$y_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = B(L)\varepsilon_t \quad (3)$$

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  $\varepsilon$  process is "white" and thus satisfies properties (2). Equation (3) says that the  $y$  process is a one-sided moving sum of a white-noise process  $\varepsilon$ .

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 *covariance generating function*  $g_y(z)$ , which is defined by

$$g_y(z) = \sum_{k=-\infty}^{\infty} c_y(k) z^k. \quad (4)$$

The coefficient on  $z^k$  in (4) is the  $k$ th lagged covariance  $c_y(k)$ .

First notice that taking mathematical expectations on both sides of (3) gives

$$E(y_t) = \sum_{j=0}^{\infty} b_j E(\varepsilon_{t-j}) = 0 \quad \text{for all } t.$$

It therefore follows that

$$c_y(k) = E\{(y_t - E y_t)(y_{t-k} - E y_{t-k})\} = E y_t y_{t-k} \quad \text{for all } k.$$

Since the  $\{\varepsilon_t\}$  process is serially uncorrelated, it follows that

$$E y_t y_{t-k} = E \left( \sum_{j=-\infty}^{\infty} b_j \varepsilon_{t-j} \cdot \sum_{h=-\infty}^{\infty} b_h \varepsilon_{t-k-h} \right) = \sigma_\varepsilon^2 \sum_{j=-\infty}^{\infty} b_j b_{j-k}$$

since only for  $j = k + h$  (or  $h = j - k$ ) is  $E \varepsilon_{t-j} \varepsilon_{t-k-h}$  nonzero and equal to  $\sigma_\varepsilon^2$ . We have permitted the  $j$  and  $h$  indexes to run over negative values, though in our

case  $b_j = 0$  for  $j < 0$ . (The formula is correct even if  $b_j \neq 0$  for  $j < 0$ .) The covariance generating function is then

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

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

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

The last equation gives the convenient expression

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

where  $B(z^{-1}) = \sum_{j=-\infty}^{\infty} b_j z^{-j}$ ,  $B(z) = \sum_{j=-\infty}^{\infty} b_j z^j$ . Equation (5) gives the covariance generating function  $g_y(z)$  in terms of the  $b_j$  and the variance  $\sigma_\varepsilon^2$  of the white noise  $\varepsilon$ .

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

$$y_t = \lambda y_{t-1} + \varepsilon_t \quad \text{or} \quad y_t = \left( \frac{1}{1 - \lambda L} \right) \varepsilon_t = \sum_{i=0}^{\infty} \lambda^i \varepsilon_{t-i}, \quad |\lambda| < 1, \quad (6)$$

where, as always,  $\varepsilon$  is a white-noise process with variance  $\sigma_\varepsilon^2$ . We have

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

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

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

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

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

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

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

$g_y(z) = \sigma_\varepsilon^2 k_1 (1 + \lambda z + \lambda^2 z^2 + \dots) + \sigma_\varepsilon^2 k_2 (z^{-1} + \lambda z^{-2} + \lambda^2 z^{-3} + \dots)$ ,  
so that  $c_y(0) = k_1 \sigma_\varepsilon^2$  and  $c_y(1) = \sigma_\varepsilon^2 \lambda k_1 = \sigma_\varepsilon^2 k_2 = c_y(-1)$ . By direct computation using (6) we note that

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

$$E y_t y_{t-1} = E \sum_{i=0}^{\infty} \lambda^i \varepsilon_{t-i} \sum_{i=1}^{\infty} \lambda^{i-1} \varepsilon_{t-i} = E \sum_{i=1}^{\infty} \lambda^i \lambda^{i-1} \varepsilon_{t-i}^2 = \sigma_\varepsilon^2 \lambda \sum_{i=1}^{\infty} \lambda^{2(i-1)} = \frac{\lambda \sigma_\varepsilon^2}{1 - \lambda^2}.$$

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

$$k_1 = \frac{1}{1 - \lambda^2}, \quad k_2 = \frac{\lambda}{1 - \lambda^2}.$$

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

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

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

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

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

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

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

Thus, we have that for the first-order Markov process (6)

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

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

$$E(y_t y_{t-k}) = \lambda E(y_{t-1} y_{t-k}) \quad \text{or} \quad c_y(k) = \lambda c_y(k-1), \quad k > 0,$$

which implies the solution

$$c_y(k) = \lambda^k c_y(0), \quad k > 0.$$

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 nonrandom part of the difference equation with initial condition  $c_y(0)$ .

As a second example, consider the second-order process

$$y_t = \left( \frac{1}{1 - \lambda_1 L} \right) \left( \frac{1}{1 - \lambda_2 L} \right) \varepsilon_t, \quad |\lambda_1 + \lambda_2| < 1, \quad \lambda_1 \neq \lambda_2, \quad (9)$$

where  $\varepsilon_t$  is white noise with variance  $\sigma_\varepsilon^2$ . Multiply both sides of (9) by  $(1 - \lambda_1 L)(1 - \lambda_2 L)$  to get

$$y_t = t_1 y_{t-1} + t_2 y_{t-2} + \varepsilon_t \quad (10)$$

where  $t_1 = \lambda_1 + \lambda_2$  and  $t_2 = -\lambda_1 \lambda_2$ . Multiply (10) by  $y_{t-k}$  for  $k > 0$  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

$$E(y_t y_{t-k}) = t_1 E(y_{t-1} y_{t-k}) + t_2 E(y_{t-2} y_{t-k}), \quad k > 0,$$

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

$$c_y(k) = t_1 c_y(k-1) + t_2 c_y(k-2). \quad (11)$$

So the covariogram of a second- (nth-) order process obeys the solution to the deterministic second- (nth-) order difference equation examined above. In particular, corresponding to (11) we consider the polynomial

$$1 - t_1 k - t_2 k^2 = 0, \quad (12)$$

which has roots  $1/\lambda_1$  and  $1/\lambda_2$ . (We know that  $1 - t_1 k - t_2 k^2$  equals  $(1 - \lambda_1 k)(1 - \lambda_2 k)$ , with roots  $1/\lambda_1$  and  $1/\lambda_2$ .) Alternatively, multiply (12) by  $k^{-2}$  to obtain

$$k^{-2} - t_1 k^{-1} - t_2 = 0, \quad x^2 - t_1 x - t_2 = 0 \quad \text{where } x = k^{-1}. \quad (13)$$

Notice that the roots of (13) are the reciprocals of the roots of (12), so  $\lambda_1$  and  $\lambda_2$  are the roots of (13).

The solution to the deterministic difference equation (11) is, as we have seen,

$$c_y(k) = \lambda_1^k z_0 + \lambda_2^k z_1, \quad k \geq 0, \quad (14)$$

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  $\lambda_1$  and  $\lambda_2$  are complex, we know from our work with deterministic difference equations and from the symmetry of covariograms that

$$c_y(k) = 2pr^k \cos(\omega k) \quad \text{or} \quad c_y(k) = c_y(0)r^k \cos(\omega k) \quad (15)$$

where  $\lambda_1 = re^{i\omega}$  and  $\lambda_2 = re^{-i\omega}$ . According to (15), the covariogram displays damped (we require  $r < 1$ ) oscillations with angular frequency  $\omega$ . A complete cycle occurs as  $\omega k$  goes from zero ( $k = 0$ ) to  $2\pi$  ( $k = 2\pi/\omega$ , 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 1 of Chapter IX.

Figure 2b displays realizations 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.

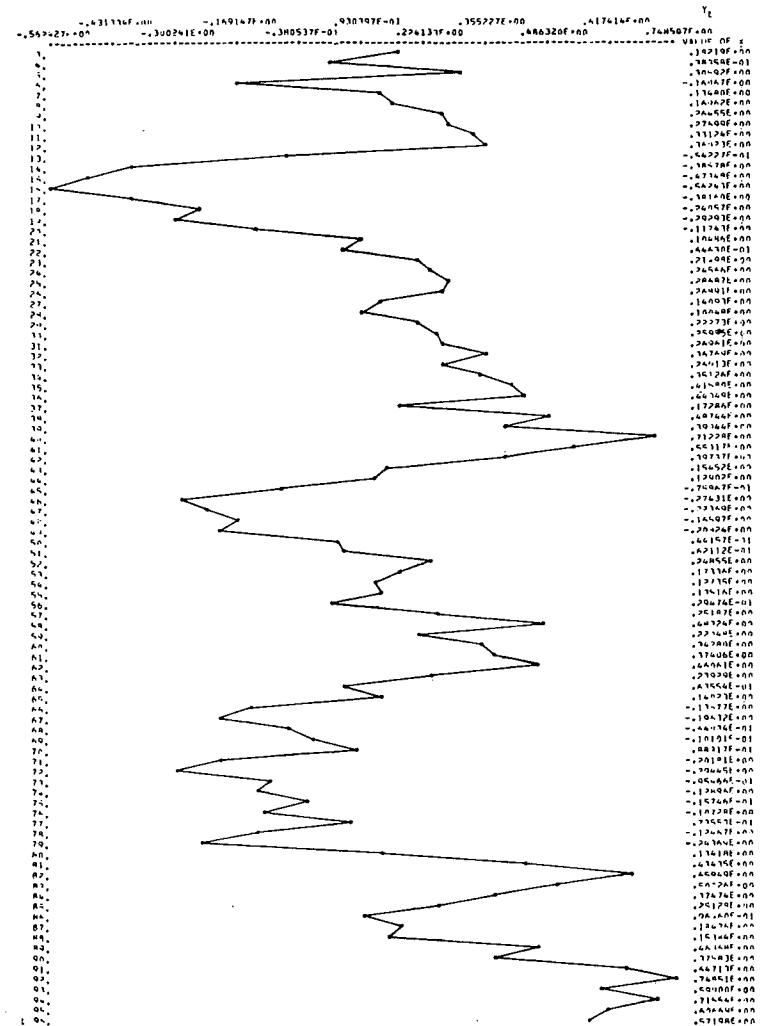
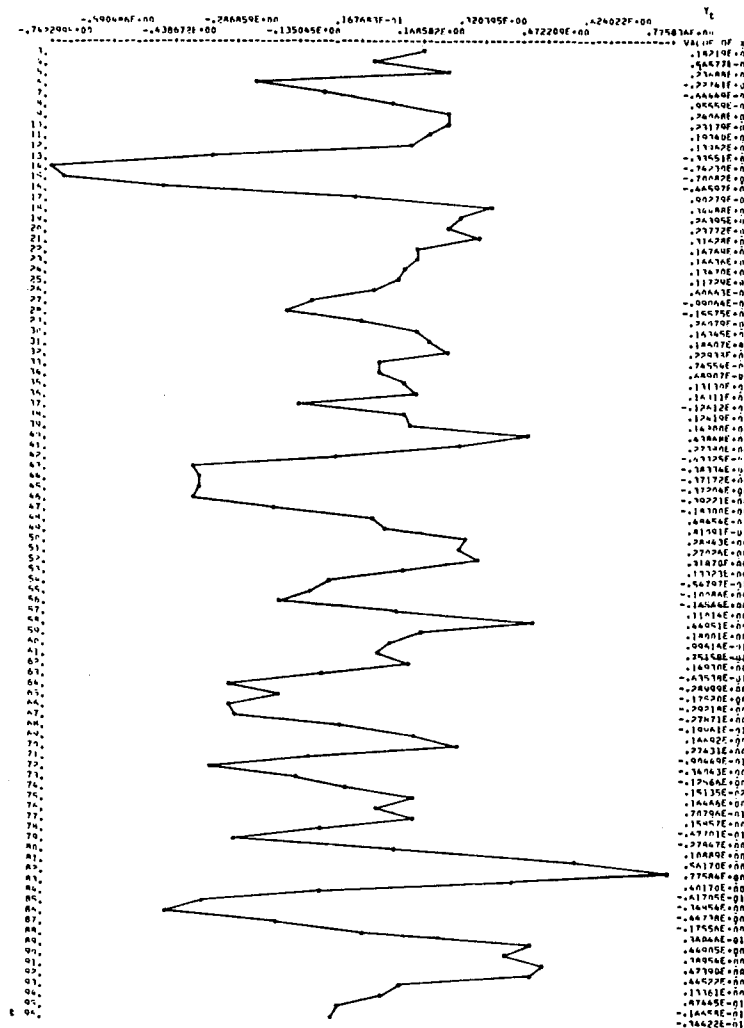
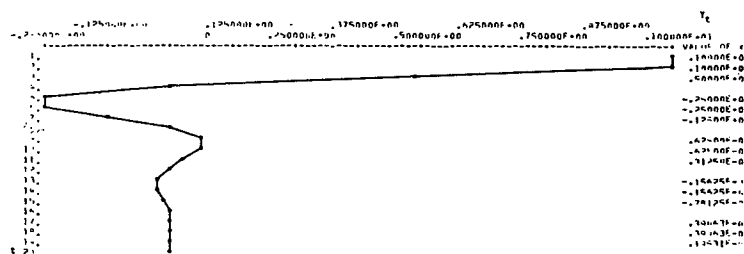


FIGURE 2a Stochastic second order system:  $Y(T) = 0.9Y(T-1) + 0Y(T-2) + \varepsilon$ .

FIGURE 2b Stochastic second order system:  $Y(T) = 1.0Y(T-1) + 0.5Y(T-2) + \varepsilon_t$ .FIGURE 2c Deterministic second order system:  $Y(T) = 1.0Y(T-1) + 0.5Y(T-2)$ .

The foregoing suggests one tentative 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  $2\pi/\omega$ , where  $\omega$  is the angular frequency associated with the damped oscillations in the covariogram (e.g., see (15)). To be labeled a business cycle the cycle should exceed a year in length. (Cycles of one year in length are termed *seasonals*.) We advance this only as a tentative definition of a cycle, and put off for a while discussing its adequacy.

### 3. THE CROSS COVARIGRAM

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

$$c_{yx}(k) = E(y_t - Ey_t)(x_{t-k} - Ex_{t-k}).$$

Now suppose that  $y_t$  and  $x_t$  can be expressed as (perhaps two-sided) distributed lags of a single white-noise process  $\varepsilon_t$ :

$$y_t = B(L)\varepsilon_t, \quad x_t = D(L)\varepsilon_t$$

where

$$B(L) = \sum_{j=-\infty}^{\infty} b_j L^j, \quad D(L) = \sum_{j=-\infty}^{\infty} d_j L^j, \quad \sum_{j=-\infty}^{\infty} b_j^2 < \infty, \quad \sum_{j=-\infty}^{\infty} d_j^2 < \infty.$$

Since  $E\varepsilon_t = 0$ , we have

$$\begin{aligned} c_{yx}(k) &= E y_t x_{t-k} = E \sum_{j=-\infty}^{\infty} b_j \varepsilon_{t-j} \sum_{h=-\infty}^{\infty} d_h \varepsilon_{t-h-k} \\ &= \sigma_\varepsilon^2 \sum_{j=-\infty}^{\infty} b_j d_{j-k}. \end{aligned}$$

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

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

In the present case, we have

$$g_{yx}(z) = \sigma_e^2 \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} b_j d_{j-k} z^k;$$

letting  $h = j - k$  so that  $k = j - h$ , we have

$$\begin{aligned} g_{yx}(z) &= \sigma_e^2 \sum_{j=-\infty}^{\infty} \sum_{h=-\infty}^{\infty} b_j d_h z^{j-h} = \sigma_e^2 \sum_{j=-\infty}^{\infty} b_j z^j \sum_{h=-\infty}^{\infty} d_h z^{-h} \\ &= \sigma_e^2 B(z)D(z^{-1}). \end{aligned} \quad (16)$$

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

Now suppose that we have the more general system

$$y_t = A(L)e_t + B(L)u_t, \quad x_t = C(L)e_t + D(L)u_t, \quad (17)$$

where  $e_t$  and  $u_t$  are two mutually uncorrelated (at all lags) white-noise processes with variances  $\sigma_e^2$  and  $\sigma_u^2$  respectively, and  $E u_t e_{t-k} = 0$  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

$$g_{yx}(z) = \sigma_e^2 A(z)C(z^{-1}) + \sigma_u^2 B(z)D(z^{-1}). \quad (18)$$

As it turns out, (17) is a very general representation for a bivariate stochastic process, including a large class of such processes.<sup>4</sup>

We define  $c_{xy}(k)$  and  $g_{xy}(z)$  symmetrically. In particular, we define

$$c_{xy}(k) = E(x_t - E x_t)(y_{t-k} - E y_{t-k}) = c_{yx}(-k).$$

We define  $g_{xy}(z)$  by

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

The particular system (17) implies that

$$g_{xy}(z) = \sigma_e^2 A(z^{-1})C(z) + \sigma_u^2 B(z^{-1})D(z).$$

#### 4. A MATHEMATICAL DIGRESSION ON FOURIER TRANSFORMS AND $z$ TRANSFORMS<sup>5</sup>

The following theorem provides the foundation for the  $z$  transform, Fourier transform, and "lag operator" methods that we use repeatedly in these pages.

<sup>4</sup> Namely, all jointly wide-sense stationary, indeterministic processes.

<sup>5</sup> This section can be omitted on first reading.

The theorem, which we shall not prove,<sup>6</sup> is a version of the Riesz-Fischer theorem.

**Theorem (Riesz-Fischer):** 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-value function  $f(\omega)$  defined for real  $\omega$ 's belonging to the interval  $[-\pi, \pi]$ , such that

$$f(\omega) = \sum_{j=-\infty}^{\infty} c_j e^{-i\omega j} \quad (19)$$

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

$$\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} \left| \sum_{j=-n}^n c_j e^{-i\omega j} - f(\omega) \right|^2 d\omega = 0.$$

The function  $f(\omega)$  is called the *Fourier transform of the  $c_k$*  and satisfies

$$\int_{-\pi}^{\pi} |f(\omega)|^2 d\omega < \infty$$

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

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) e^{+i\omega k} d\omega. \quad (20)$$

Finally, the function  $f(\omega)$  and the  $c_k$  satisfy Parseval's relation

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

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  $l_2(-\infty, \infty)$ . It is a *linear space* in the sense that it possesses the following two properties (among others)<sup>7</sup>:

(i) Let  $\alpha$  be a scalar and let  $\{x_t\}$  belong to  $l_2(-\infty, \infty)$ . Then  $\{\alpha x_k\}$  belongs to  $l_2(-\infty, \infty)$ , i.e.,  $\sum_{k=-\infty}^{\infty} |\alpha x_k|^2 < \infty$ .

(ii) Let  $\{x_k\}$  and  $\{y_k\}$  both belong to  $l_2(-\infty, \infty)$ . Then  $\{x_k + y_k\}$  belongs to  $l_2(-\infty, \infty)$ , i.e.,  $\sum_{k=-\infty}^{\infty} |x_k + y_k|^2 < \infty$ .

Now consider the space  $L_2[-\pi, \pi]$  consisting of all functions  $f(\omega)$  for which  $\int_{-\pi}^{\pi} |f(\omega)|^2 d\omega < \infty$ , i.e., the space of "square Lebesgue integrable functions"

<sup>6</sup> For a proof of the Riesz-Fischer theorem, see Apostol (1974, Chapter 11).

<sup>7</sup> For a statement of the defining properties of linear spaces, see Naylor and Sell (1971).



on  $[-\pi, \pi]$ . We denote this space  $L_2[-\pi, \pi]$ . This space is a *linear space* in the sense that it possesses the two properties (among some others):

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

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

The spaces  $l_2(-\infty, \infty)$  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  $l_2(-\infty, \infty)$  the real-valued function

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

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  $l_2(-\infty, \infty)$  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

$$D_2(f, g) = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\omega) - g(\omega)|^2 d\omega \right\}^{1/2}$$

is a metric<sup>8</sup> that measures the "distance" between two functions  $f(\omega)$  and  $g(\omega)$ . The metric  $D_2(\cdot, \cdot)$  is defined for all  $f(\omega)$  and  $g(\omega)$  belonging to  $L_2[-\pi, \pi]$ .

Now consider the mapping from  $l_2(-\infty, \infty)$  to  $L_2[-\pi, \pi]$  defined by the Fourier transform

$$f(\omega) = \sum_{k=-\infty}^{\infty} c_k e^{-i\omega k}, \quad \omega \in [-\pi, \pi]. \quad (19)$$

We also have the inverse mapping

$$c_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) e^{+i\omega j} d\omega, \quad j = 0, \pm 1, \pm 2, \dots \quad (20)$$

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

$$f(\omega) = \sum_{k=-\infty}^{\infty} c_k e^{-i\omega k}$$

where

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) e^{+i\omega k} d\omega$$

<sup>8</sup> We adopt the usual convention that if  $f = g$  except on a set of Lebesgue measure zero, we agree to say that the functions  $f$  and  $g$  are equal. On this convention  $D_2(f, g)$  is a metric. See Naylor and Sell (1971) for more details.

and where the infinite sum converges in the mean square sense. This converse theorem assures us that the mapping of  $l_2(-\infty, \infty)$  into  $L_2[-\pi, \pi]$  defined by (19) is *onto*. It is also one-to-one. The usefulness of the mapping (19) stems from the fact that it is an *isometric isomorphism* from  $l_2(-\infty, \infty)$  to  $L_2[-\pi, \pi]$ ; i.e., it is a *one-to-one* and *onto* transformation of points in  $l_2(-\infty, \infty)$  into points in  $L_2[-\pi, \pi]$  that preserves both linear structure (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  $l_2(-\infty, \infty)$ , let  $\alpha$  be a scalar, and let

$$x(\omega) = \sum_{k=-\infty}^{\infty} x_k e^{-i\omega k}, \quad y(\omega) = \sum_{k=-\infty}^{\infty} y_k e^{-i\omega k}.$$

Then we have (as can be verified directly)

$$x(\omega) + y(\omega) = \sum_{k=-\infty}^{\infty} (x_k + y_k) e^{-i\omega k}, \quad \alpha x(\omega) = \sum_{k=-\infty}^{\infty} \alpha x_k e^{-i\omega k}.$$

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  $\alpha$  times the Fourier transform of  $\{x_k\}$ ." This means that (19) is an isomorphism. We also have

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

or

$$D_2(x(\omega), y(\omega)) = d_2(x, y),$$

so that (19) is an isometric mapping.

The Fourier transformation (19) puts square summable sequences  $\{x_k\}$  into one-to-one correspondence with square integrable functions  $f(\omega)$  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 *convolution* of two sequences  $\{x_k\}$  and  $\{y_k\}$  defined to be the new sequence

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

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

$$\sum_{k=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} y_s x_{k-s} e^{-i\omega k} = \sum_{s=-\infty}^{\infty} y_s e^{-i\omega s} \sum_{k=-\infty}^{\infty} x_{k-s} e^{-i\omega(k-s)} = y(\omega)x(\omega)$$

where  $y(\omega) = \sum_{k=-\infty}^{\infty} y_k e^{-i\omega k}$ ,  $x(\omega) = \sum_{k=-\infty}^{\infty} x_k e^{-i\omega k}$ . Thus the Fourier transform of the *convolution* 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 (20) after the calculations have been performed.

By making the change of variable  $z = e^{i\omega}$  in the Riesz–Fischer 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 plane such that

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

where the infinite series converges in the mean square sense that

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

where  $\Gamma$  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

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

The function  $g(z)$  is called the  $z$  transform of the sequence  $\{c_k\}$ . The  $c_k$  can be recovered from  $g(z)$  by  $c_k = (2\pi i)^{-1} \int_{\Gamma} g(z) z^{-k-1} dz$ . 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  $e^{i\omega j}$ ,  $j = 0, \pm 1, \pm 2, \dots$  are an orthogonal set on the interval  $[-\pi, \pi]$ . That is, for  $n \neq m$ , we have

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega n} e^{-i\omega m} d\omega &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega(n-m)} d\omega = \frac{1}{2\pi i(n-m)} [e^{i\omega(n-m)}]_{-\pi}^{\pi} \\ &= \frac{1}{2\pi i(n-m)} [e^{i\pi(n-m)} - e^{-i\pi(n-m)}] \\ &= \frac{1}{\pi(n-m)} \sin \pi(n-m) = 0 \end{aligned}$$

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

For the most part, the Riesz–Fischer theorem and its corollary are sufficient for our needs. Below we shall briefly touch on a deterministic process for which the condition  $\sum |c_k|^2 < \infty$  is violated (where the  $c_k$  depict the covariogram) so that the theorem will not suffice to define the Fourier transform of the  $c_k$ . 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.

## 5. 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 (4),

$$g_y(z) = \sum_{k=-\infty}^{\infty} c_y(k) z^k. \quad (4)$$

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

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

If we evaluate (4) at the value  $z = e^{-i\omega}$ , we have

$$g_y(e^{-i\omega}) = \sum_{k=-\infty}^{\infty} c_y(k) e^{-i\omega k}, \quad -\pi < \omega < \pi. \quad (21)$$

Viewed as a function of angular frequency  $\omega$ ,  $g_y(e^{-i\omega})$  is called the *spectrum* of  $y$ . The spectrum is the Fourier transform of the covariogram.

As we would expect from the inversion formula (20), the spectrum is itself a kind of covariance generating function. Given an expression for  $g_y(e^{-i\omega})$  it is easy to recover the covariances  $c_y(k)$  from the inversion formula (20). To motivate the inversion formula, we multiply (21) by  $e^{i\omega h}$  and integrate with respect to  $\omega$  from  $-\pi$  to  $\pi$ :

$$\int_{-\pi}^{\pi} g_y(e^{-i\omega}) e^{i\omega h} d\omega = \int_{-\pi}^{\pi} \sum_{k=-\infty}^{\infty} c_y(k) e^{i\omega(h-k)} d\omega = \sum_{k=-\infty}^{\infty} c_y(k) \int_{-\pi}^{\pi} e^{i\omega(h-k)} d\omega. \quad (22)$$

Now for  $h = k$ , we have

$$\int_{-\pi}^{\pi} e^{i\omega(h-k)} d\omega = \int_{-\pi}^{\pi} 1 d\omega = 2\pi.$$

For  $h \neq k$ , we have

$$\begin{aligned} \int_{-\pi}^{\pi} e^{i\omega(h-k)} d\omega &= \int_{-\pi}^{\pi} \cos \omega(h-k) d\omega + i \int_{-\pi}^{\pi} \sin \omega(h-k) d\omega \\ &= -\sin \omega(h-k) \Big|_{-\pi}^{\pi} + i \cos \omega(h-k) \Big|_{-\pi}^{\pi} = 0. \end{aligned}$$

Therefore, (22) becomes

$$\int_{-\pi}^{\pi} g_y(e^{-i\omega}) e^{i\omega h} d\omega = 2\pi c_y(h).$$

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

$$\int_{-\pi}^{\pi} g_y(e^{-i\omega}) d\omega = 2\pi c_y(0),$$

so that the area under the spectrum from  $-\pi$  to  $\pi$  equals  $2\pi$  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 (21) we have

$$\begin{aligned} g_y(e^{-i\omega}) &= \sum_{k=-\infty}^{\infty} c_y(k) e^{-i\omega k} \\ &= c_y(0) + \sum_{k=1}^{\infty} c_y(k) (e^{i\omega k} + e^{-i\omega k}) \\ &= c_y(0) + 2 \sum_{k=1}^{\infty} c_y(k) \cos \omega k. \end{aligned} \quad (23)$$

According to (23) 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 (23) that

$$g_y(e^{i\omega}) = g_y(e^{-i\omega}),$$

so that the spectrum is symmetric about  $\omega = 0$ .

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

We now derive a fundamental formula linking the spectrum of one covariance stationary process  $y_t$  to the spectrum of another covariance stationary process  $x_t$ . We suppose that both  $\{x_t\}$  and  $\{y_t\}$  have zero mean and consider the projection equation

$$y_t = \sum_{j=-\infty}^{\infty} b_j x_{t-j} + \varepsilon_t \equiv B(L)x_t + \varepsilon_t$$

where  $E\varepsilon_t x_{t-j} = 0$  for all  $j$ . Here  $B(L)x_t$  is the projection of  $y_t$  on the entire  $x$  process, as is implied by the orthogonality principle. We then have that

$$\begin{aligned} y_t y_{t-j} &= \left( \sum_{s=-\infty}^{\infty} b_s x_{t-s} \right) \left( \sum_{r=-\infty}^{\infty} b_r x_{t-j-r} \right) \\ &+ \left( \sum_{s=-\infty}^{\infty} b_s x_{t-s} \right) \varepsilon_{t-j} + \left( \sum_{r=-\infty}^{\infty} b_r x_{t-j-r} \right) \varepsilon_t + \varepsilon_t \varepsilon_{t-j}. \end{aligned}$$

Taking expected values of both sides and applying the orthogonality conditions gives

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

The spectrum of  $y$  is defined as

$$\begin{aligned} g_y(e^{-i\omega}) &= \sum_{k=-\infty}^{\infty} c_y(k) e^{-i\omega k} \\ &= \sum_{k=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} b_r b_s c_x(k+r-s) e^{-i\omega k} + g_\varepsilon(e^{-i\omega}). \end{aligned} \quad (24)$$

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

$$e^{-i\omega k} = e^{-i\omega(h-r+s)} = e^{-i\omega h} e^{i\omega r} e^{-i\omega s}. \quad (25)$$

Substituting (25) into (24) gives

$$\begin{aligned} g_y(e^{-i\omega}) &= \sum_{r=-\infty}^{\infty} b_r e^{i\omega r} \sum_{s=-\infty}^{\infty} b_s e^{-i\omega s} \sum_{h=-\infty}^{\infty} c_x(h) e^{-i\omega h} + g_\varepsilon(e^{-i\omega}) \\ &= B(e^{i\omega}) B(e^{-i\omega}) g_x(e^{-i\omega}) + g_\varepsilon(e^{-i\omega}) \end{aligned}$$

or

$$g_y(e^{-i\omega}) = |B(e^{-i\omega})|^2 g_x(e^{-i\omega}) + g_\varepsilon(e^{-i\omega}). \quad (26)$$

This is an important formula that shows how the spectrum of the "input"  $x$  is multiplied by the nonnegative real number  $|B(e^{-i\omega})|^2$  in composing the spectrum of  $y$ .

Formula (26) can be used to analyze the effects of "filtering," in which we start with a covariance stationary random process  $x_t$  and define a new process

$$y_t = B(L)x_t, \quad (27)$$

so that formula (26) applies with  $g_\varepsilon(e^{-i\omega}) \equiv 0$ . We shall illustrate the usefulness of this formula in several contexts. To begin, formula (27) motivates the interpretation of the spectrum as decomposing the variance of  $y$  by frequency. Thus, suppose we could choose  $B(e^{-i\omega})$  so that

$$B(e^{-i\omega}) = \begin{cases} 1 & \text{for } \omega \in [a, b] \cup [-b, -a], \quad 0 < a < b < \pi, \\ 0 & \text{otherwise.} \end{cases} \quad (28)$$

Thus, we are choosing a "filter," i.e., a set of  $b_j$ , that takes a random process  $x_t$  and transforms it into a random process  $y_t$  according to (27). A filter obeying (28) shuts off all of the spectral power for frequencies not in the region  $[a, b]$  or  $[-b, -a]$ . To determine a set of  $b_j$  that satisfies (28), we use the "inversion" formula seen earlier,

$$\begin{aligned} b_j &= \frac{1}{2\pi} \int_{-\pi}^{\pi} B(e^{-i\omega}) e^{+i\omega j} d\omega = \frac{1}{2\pi} \int_{-b}^{-a} e^{i\omega j} d\omega + \frac{1}{2\pi} \int_a^b e^{i\omega j} d\omega \\ &= \frac{1}{2\pi} \int_a^b (e^{i\omega j} + e^{-i\omega j}) d\omega = \frac{1}{2\pi} \int_a^b 2 \cos \omega j d\omega \\ &= \frac{1}{\pi} \frac{1}{j} \sin \omega j \Big|_a^b \\ &= \frac{1}{\pi} \left( \frac{\sin jb - \sin ja}{j} \right), \quad \text{for all integers } j. \end{aligned} \quad (29)$$

Note that  $b_j = b_{-j}$ . With the  $b_j$  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  $\omega \in [a, b]$ ,  $\omega \in [-b, -a]$ . The variance of  $y$  is given by

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} g_y(e^{-i\omega}) d\omega = \frac{1}{2\pi} \int_{-b}^{-a} g_x(e^{-i\omega}) d\omega + \frac{1}{2\pi} \int_a^b g_x(e^{-i\omega}) d\omega.$$

In this sense  $g_x(e^{-i\omega})$  gives a decomposition of the variance of  $x$  by frequency, the variance occurring over a given frequency band being found by integrating the spectrum over that band and dividing by  $2\pi$ . We have already seen that by integrating the spectrum from  $-\pi$  to  $\pi$  we obtain the variance of  $x$  times  $2\pi$ . As we shall show presently, 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 (28) 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  $g_x(e^{-i\omega})$  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  $y_t = \varepsilon_t$ ,  $\varepsilon_t$  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

$$g_y(e^{-i\omega}) = \sigma_\varepsilon^2, \quad -\pi \leq \omega \leq \pi.$$

The spectrum is flat, and equals  $\sigma_\varepsilon^2$  at each frequency. Notice that

$$\int_{-\pi}^{\pi} g_y(e^{-i\omega}) d\omega = 2\pi\sigma_\varepsilon^2,$$

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

Next consider the first-order process

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

For this process the covariance generating function is

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

Therefore, the spectrum is

$$\begin{aligned} g_y(e^{-i\omega}) &= \left( \frac{1}{1 - \lambda e^{-i\omega}} \right) \left( \frac{1}{1 - \lambda e^{i\omega}} \right) \sigma_\varepsilon^2 = \frac{1}{1 - \lambda(e^{i\omega} + e^{-i\omega}) + \lambda^2} \sigma_\varepsilon^2 \\ &= \frac{1}{1 - 2\lambda \cos \omega + \lambda^2} \sigma_\varepsilon^2 \end{aligned}$$

Notice that

$$\frac{dg_y(e^{-i\omega})}{d\omega} = -(1 - 2\lambda \cos \omega + \lambda^2)^{-2} (2\lambda \sin \omega) \sigma_\varepsilon^2.$$

The first term in parenthesis is positive. Since  $\sin \omega > 0$  for  $0 < \omega < \pi$ , 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  $\omega$  increases; if  $\lambda < 0$ , the spectrum increases on  $(0, \pi)$  as  $\omega$  increases. Thus, if  $\lambda > 0$ , low frequencies (i.e., low values of  $\omega$ ) are relatively important in composing the variance of  $\omega$ , while if  $\lambda < 0$ , high frequencies are the more important. It is easy to verify that the higher in absolute value is  $\lambda$ , the steeper is the spectrum. Notice that the first-order process can have a peak in its spectrum only at  $\omega = 0$  or  $\omega = \pm\pi$ . A peak at  $\omega = \pi$  corresponds to a periodicity of  $2\pi/\omega = 2\pi/\pi = 2$  periods. A peak at  $\omega = 0$  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.

A second-order process can have a peak in its spectrum inside the interval  $(0, \pi)$ . Consider the second-order process

$$y_t = \frac{1}{1 - t_1 L - t_2 L^2} \varepsilon_t,$$

$\varepsilon_t$  white noise. For this process, the covariance generating function is

$$g_y(z) = \left( \frac{1}{1 - t_1 z - t_2 z^2} \right) \left( \frac{1}{1 - t_1 z^{-1} - t_2 z^{-2}} \right) \sigma_\varepsilon^2.$$

Therefore, the spectrum of the process is

$$\begin{aligned} g_y(e^{-i\omega}) &= \left( \frac{1}{1 - t_1 e^{-i\omega} - t_2 e^{-2i\omega}} \right) \left( \frac{1}{1 - t_1 e^{i\omega} - t_2 e^{2i\omega}} \right) \sigma_\varepsilon^2 \\ &= \frac{\sigma_\varepsilon^2}{1 + t_1^2 + t_2^2 + (t_2 t_1 - t_1)(e^{i\omega} + e^{-i\omega}) - t_2(e^{-2i\omega} + e^{2i\omega})} \\ &= \frac{\sigma_\varepsilon^2}{1 + t_1^2 + t_2^2 - 2t_1(1 - t_2) \cos \omega - 2t_2 \cos 2\omega} = \frac{\sigma_\varepsilon^2}{h(\omega)}. \end{aligned}$$

Differentiating with respect to  $\omega$ , we have

$$\begin{aligned} \frac{dg_y(e^{-i\omega})}{d\omega} &= -\sigma_\varepsilon^2 h(\omega)^{-2} (2t_1(1 - t_2) \sin \omega + 4t_2 \sin 2\omega) \\ &= -\sigma_\varepsilon^2 h(\omega)^{-2} ((2 \sin \omega)[t_1(1 - t_2) + 4t_2 \cos \omega]). \end{aligned}$$

We know that  $h(\omega)^2 > 0$ . For the above derivative to be zero at a  $\omega$  belonging to  $(0, \pi)$ , we must have the term in brackets equal to zero:

$$t_1(1 - t_2) + 4t_2 \cos \omega = 0 \quad \text{or} \quad \cos \omega = \frac{-t_1(1 - t_2)}{4t_2}, \quad (30)$$

so that

$$\omega = \cos^{-1} \left( \frac{-t_1(1 - t_2)}{4t_2} \right). \quad (31)$$

Equation (31) can be satisfied only if

$$\left| \frac{-t_1(1 - t_2)}{4t_2} \right| < 1 \quad (32)$$

since  $|\cos x| \leq 1$  for all  $x$ . By inspecting the second derivative of  $g_y(e^{-i\omega})$  with respect to  $\omega$ , it can be verified that at the  $\omega$  given by (31) there is a peak in the spectrum if  $t_2 < 0$  and a trough if  $t_2 > 0$ . Condition (32) is slightly more restrictive than the condition that the roots of the deterministic difference equation be complex so that the covariogram displays oscillations. Let us write (32) as

$$-1 < -t_1(1 - t_2)/4t_2 < 1. \quad (33)$$

The boundaries of the region (33) are

$$-t_1(1 - t_2) = 4t_2 \quad (34)$$

and

$$-t_1(1 - t_2) = -4t_2. \quad (35)$$

The points  $(t_1, t_2) = (0, 0)$  appear on both boundaries, while the point  $(t_1, t_2) = (2, -1)$  appears on (34) and  $(t_1, t_2) = (-2, -1)$  appears on (35). Differentiating (34) implicitly with respect to  $t_1$  gives

$$dt_2/dt_1 = (t_2 - 1)/(4 - t_1)$$

so that along (34)

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

and

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

Differentiating (35) with respect to  $t_1$  gives

$$dt_2/dt_1 = (1 - t_2)/(4 + t_1)$$

so that along (35)

$$\left. \frac{dt_2}{dt_1} \right|_{t_1=t_2=0} = \frac{1}{4}, \quad \left. \frac{dt_2}{dt_1} \right|_{t_1=-2, t_2=-1} = 1.$$

Such calculations show that the boundaries of region (34) are as depicted in Figure 3. To be in region (33) with  $t_2 < 1$  (a requirement of covariance stationarity) implies that the roots of the difference equation are complex. However, complex roots do not imply that (33) is satisfied. Consequently, the conditions for an oscillatory covariogram are not quite equivalent with those for a spectral peak.

To illustrate the ability of low-order stochastic difference equations to generate "realistic" data, Figures 2a and 2b show simulations of first- and second-order stochastic difference equations, while Figure 2c shows the solution

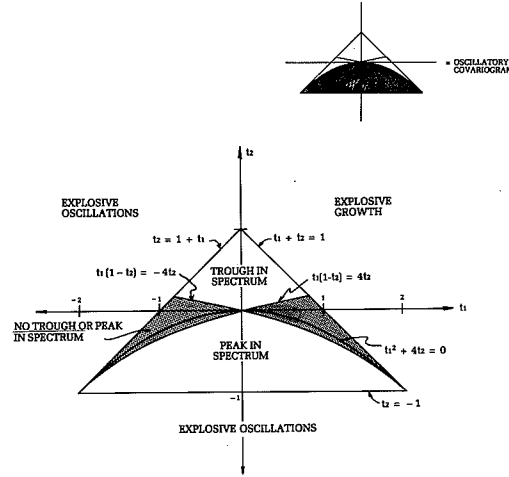


FIGURE 3 (Source: G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*, p. 229, San Francisco: Holden-Day, 1969.)

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

$$y_t = 0.9y_{t-1} + \varepsilon_t,$$

$\varepsilon_t$ , 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 do not have peaks on  $(0, \pi)$  and even if their covariograms do not oscillate.

## 6. 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

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

## 6. THE CROSS SPECTRUM

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

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

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

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

$$c_{yx}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g_{yx}(e^{-i\omega})e^{i\omega k} d\omega.$$

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

$$g_{xy}(e^{-i\omega}) = \overline{g_{yx}(e^{-i\omega})} = g_{yx}(e^{+i\omega}) \quad (36)$$

where the bar denotes complex conjugation and

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

and  $c_{xy}(k) = E x_t y_{t-k}$ . 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  $\varepsilon_t$  by

$$y_t = \sum_{j=-\infty}^{\infty} h_j x_{t-j} + \varepsilon_t \quad (37)$$

where  $E\varepsilon_t = Ex_t = 0$ , and  $E\varepsilon_t x_{t-s} = 0$  for all  $s$ , an orthogonality condition that characterizes  $\sum h_j x_{t-j}$  as the projection of  $y_t$  on the space spanned by  $\{x_{t-\infty}, \dots, x_0, \dots, x_{t+\infty}\}$ . Then we have already seen that the spectrum of  $y$  satisfies

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

where

$$h(e^{-i\omega}) = \sum_{j=-\infty}^{\infty} h_j e^{-i\omega j}.$$

To find the cross spectrum between  $y$  and  $x$ , first use (37) to calculate the  $k$ th lagged covariance as

$$E y_t x_{t-k} = \sum_{j=-\infty}^{\infty} h_j E(x_{t-j} x_{t-k}), \quad c_{yx}(k) = \sum_{j=-\infty}^{\infty} h_j c_x(k-j).$$

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

$$g_{yx}(e^{-i\omega}) = h(e^{-i\omega})g_x(e^{-i\omega})$$

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  $e^{-i\omega k}$  and summing over  $k$ ) gives

$$\sum_{k=-\infty}^{\infty} c_{yx}(k) e^{-i\omega k} = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} h_j c_x(k-j) e^{-i\omega k}.$$

Noting that  $e^{-i\omega k} = e^{-i\omega(k-j)} e^{-i\omega j}$ , the above can be written as

$$g_{yx}(e^{-i\omega}) = \sum_{j=-\infty}^{\infty} h_j e^{-i\omega j} \sum_{k=-\infty}^{\infty} c_x(k-j) e^{-i\omega(k-j)}$$

or

$$g_{yx}(e^{-i\omega}) = h(e^{-i\omega})g_x(e^{-i\omega}). \quad (38)$$

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

$$c_{yx}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} h(e^{-i\omega})g_x(e^{-i\omega})e^{i\omega k} d\omega.$$

Further, notice that given  $g_{yx}(e^{-i\omega})$  and  $g_x(e^{-i\omega})$ , the  $h_k$  can be recovered from

$$h_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{g_{yx}(e^{-i\omega})}{g_x(e^{-i\omega})} e^{i\omega k} d\omega.$$

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

As an example, suppose that the jointly covariance stationary process  $(y, x)$  has covariance generating functions

$$g_x(z) = \sigma_e^2 \left( \frac{1}{1-0.9z} \right) \left( \frac{1}{1-0.9z^{-1}} \right), \quad g_y(z) = \sigma_u^2 (1-0.8z)(1-0.8z^{-1})$$

$$g_{yx}(z) = \sigma_{ue} (1-0.8z)(1+0.5z^{-1}).$$

Notice that this is equivalent with

$$g_x(e^{-i\omega}) = \sigma_e^2 \frac{1}{(1-0.9e^{-i\omega})(1-0.9e^{+i\omega})} = \frac{\sigma_e^2}{1.81-1.8\cos\omega}$$

$$g_y(e^{-i\omega}) = \sigma_u^2 (1.64-1.6\cos\omega)$$

$$g_{yx}(e^{-i\omega}) = \sigma_{ue} (0.6-0.8e^{-i\omega}+0.5e^{+i\omega}).$$

Let us now use formula (38) to calculate the coefficient generating function  $h(z)$  in the projection of  $y_t$  on the entire  $x$  process. Using  $z$  instead of  $e^{-i\omega}$ , formula (38) becomes

$$h(z) = g_{yx}(z)/g_x(z).$$

For our example this gives

$$g_{yx}(z)/g_x(z) = (\sigma_{ue}/\sigma_e^2)(1-0.8z)(1+0.5z^{-1})(1-0.9z^{-1})(1-0.9z).$$

The reader can easily multiply this polynomial in  $z$  and verify that  $h_j = 0$  for  $|j| > 2$  and that  $h_j \neq 0$  for  $j = -2, -1, 0, 1, 2$ . Notice that, as in general,  $h(z)$  is "two-sided," having nonzero coefficients on negative powers of  $z$ .

Now let us calculate the coefficients in the projection of  $x_t$  on the entire  $y$  process:

$$x_t = \sum_{j=-\infty}^{\infty} f_j y_{t-j} + u_t$$

$E u_t y_{t-j} = 0$  for all  $j$ . Applying formula (38), exchanging the roles of  $y$  and  $x$ , gives

$$f(z) = g_{xy}(z)/g_y(z) = g_{yx}(z^{-1})/g_y(z).$$

In our example this gives

$$f(z) = \frac{\sigma_{ue}(1-0.8z^{-1})(1+0.5z)}{\sigma_u^2(1-0.8z^{-1})(1-0.8z)} = \frac{\sigma_{ue}}{\sigma_u^2} \frac{(1+0.5z)}{(1-0.8z)}.$$

It is readily verified that  $f_j = 0$  for  $j < 0$ , so that  $f(z)$  is "one-sided on the past and present."

We shall shortly study the conditions under which the projection of  $y$  on  $x$  or of  $x$  on  $y$  are "one-sided" or "two-sided."

Equation (16) can be generalized as follows. Let

$$y_{1t} = B_1(L)x_t, \quad y_{2t} = B_2(L)x_t \quad (39)$$

where  $x_t$  is a covariance stationary process and  $B_1(L)$  and  $B_2(L)$  are the lag generating functions for square summable lag distributions. Then

$$g_{y_1 y_2}(e^{-i\omega}) = B_1(e^{-i\omega})B_2(e^{+i\omega})g_x(e^{-i\omega}). \quad (40)$$

We invite the reader to verify this formula by using (39) to calculate  $Ey_{1t}y_{2t-k}$ , multiplying by  $e^{-i\omega k}$ , and summing over  $k$ . The derivation mimics the derivation of Equation (16) above.<sup>9</sup>

We now use formula (40) to show that the spectrum reflects a decomposition of  $x_t$  into processes that are orthogonal across frequencies. Thus let

$$y_{1t} = B_1(L)x_t, \quad y_{2t} = B_2(L)x_t$$

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

$$B_1(e^{-i\omega}) = \begin{cases} 1, & \omega \in [-b, -a] \cup [a, b], \\ 0, & \omega \notin [-b, -a] \cup [a, b]; \end{cases}$$

$$B_2(e^{-i\omega}) = \begin{cases} 1, & \omega \in [-d, -c] \cup [c, d], \\ 0, & \omega \notin [-d, -c] \cup [c, d]. \end{cases}$$

To find the individual distributed lag coefficients, Equation (29) can be used. Equation (40) evaluated at  $z = e^{-i\omega}$  implies

$$g_{y_1 y_2}(e^{-i\omega}) = B_1(e^{-i\omega})B_2(e^{i\omega})g_x(e^{-i\omega}).$$

If  $[-b, -a] \cup [a, b]$  does not intersect the set of frequencies  $[-d, -c] \cup [c, d]$ , then  $B_1(e^{-i\omega})B_2(e^{i\omega}) = 0$  for all  $\omega$ , so that  $g_{y_1 y_2}(e^{-i\omega}) = 0$ . 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  $g_x(e^{-i\omega})$  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

$$g_{yx}(e^{-i\omega}) = \text{co}(\omega) + i \text{qu}(\omega)$$

where  $\text{co}(\omega)$  is called the cospectrum and  $\text{qu}(\omega)$  is called the quadrature spectrum. A more usual representation is the polar one

$$g_{yx}(e^{-i\omega}) = r(\omega)e^{i\theta(\omega)} \quad (41)$$

where

$$r(\omega) = \sqrt{\text{co}(\omega)^2 + \text{qu}(\omega)^2}, \quad \theta(\omega) = \tan^{-1} \left[ \frac{\text{qu}(\omega)}{\text{co}(\omega)} \right].$$

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

$$\text{coh}(\omega) = |g_{yx}(e^{-i\omega})|^2 / g_x(e^{-i\omega})g_y(e^{-i\omega}),$$

<sup>9</sup> Alternatively, write  $x_t$  in terms of its moving average representation  $x_t = c(L)\eta_t$  where  $g_x(z) = \sigma_\eta^2 c(z)c(z^{-1})$ . Then apply (16) to the system  $y_{1t} = B_1(L)c(L)\eta_t$ ,  $y_{2t} = B_2(L)c(L)\eta_t$ .

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  $\omega$  that is accounted for by variation in the other series.

Notice that from (38) and from the fact that the spectrum  $g_x(e^{-i\omega})$  is real, the phase of the cross spectrum equals the phase of  $h(e^{-i\omega}) = \sum h_j e^{-i\omega j}$ , which is the Fourier transform of the  $h_j$ . That is, writing (38) and (41), we have

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

or

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

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

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

where  $s(\omega) = r(\omega)/g_x(e^{-i\omega})$ .

The following provides a heuristic device for interpreting  $\theta(\omega)$ . Suppose we consider as an input into the system (37) an  $x$  series consisting of a pure cosine wave of frequency  $\omega$ :

$$x_t = 2 \cos \omega t = e^{i\omega t} + e^{-i\omega t}.$$

For this input path, suppressing the disturbance  $\varepsilon_t$ , (37) becomes

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

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

$$y_t = e^{i\omega t} s(\omega)e^{i\theta(\omega)} + e^{-i\omega t} s(\omega)e^{-i\theta(\omega)} = s(\omega)[e^{i(\omega t + \theta(\omega))} + e^{-i(\omega t + \theta(\omega))}]$$

$$= s(\omega)2 \cos(\omega t + \theta(\omega)).$$

Therefore, the response of (37) to an input in the form of a cosine wave of frequency  $\omega$  is a cosine wave at the same frequency with amplitude multiplied by  $s(\omega)$  and phase shifted by  $\theta(\omega)$ . The input cosine wave is at its peak at  $t = 0$ , while the output is at its peak at  $\omega t + \theta(\omega) = 0$  or  $t = -\theta(\omega)/\omega$  units of time. Thus, for  $\theta(\omega) > 0$ , the output *leads* the input by  $-\theta(\omega)/\omega$  units of time (where we adopt the usual convention that  $\theta(\omega)$  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 non-negligible band of frequencies that there occurs a positive contribution to variance. Thus, for such processes, there really do not 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 are not generated by the stochastic difference equations that we are studying. (See pp. 260–262.)

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

$$h(e^{-i\omega}) = \sum_j h_j e^{-i\omega j},$$

we note that  $h(e^{-i\omega})$  evaluated at  $\omega = 0$  is the sum of the lag weights, i.e.,

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

Notice that since

$$\sum h_j e^{-i\omega j} = \sum h_j \cos \omega j - i \sum h_j \sin \omega j$$

and that since  $\sin 0 = 0$ , we have that

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

Since  $h(e^{-i\omega})$  is real at zero frequency, the phase statistic  $\theta(\omega)$  is zero at zero frequency, provided  $\sum h_j \neq 0$ :

$$\theta(\omega) = \tan^{-1} \left[ -\frac{\sum h_j \sin \omega j}{\sum h_j \cos \omega j} \right], \quad \theta(0) = \tan^{-1}[0] = 0. \quad (42)$$

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

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

Applying this to (42) gives

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

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

$$\theta'(0) = -\sum h_j j / \sum h_j.$$

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

## 7. 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.<sup>10</sup> 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, such as 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 in terms of categorizing 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

$$x_t = \lambda x_{t-1} + u_t, \quad |\lambda| < 1, \quad y_t = h_0 x_t + h_1 x_{t-1} + \varepsilon_t \quad (43)$$

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

$$g_{yx}(e^{-i\omega}) = (h_0 + h_1 e^{-i\omega}) g_x(e^{-i\omega}) = (h_0 + h_1 \cos \omega - i h_1 \sin \omega) g_x(e^{-i\omega}) \\ = r(\omega) e^{i\theta(\omega)} g_x(e^{-i\omega})$$

where

$$r(\omega) = \sqrt{(h_0 + h_1 \cos \omega)^2 + (h_1 \sin \omega)^2}, \quad \theta(\omega) = \tan^{-1} \left[ \frac{-h_1 \sin \omega}{h_0 + h_1 \cos \omega} \right].$$

Now by suitably choosing  $h_0$  and  $h_1$ , at a given frequency  $\theta(\omega)$  can be set arbitrarily in the interval  $(-\pi, \pi)$ . This is in spite of the fact that the model (43) 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

$$x_t = \lambda x_{t-1} + u_t, \quad y_t = (x_t - x_{t-1}) + \varepsilon_t,$$

<sup>10</sup> Leading indicators are published in *Business Conditions Digest*, published by the Department of Commerce.

where as before  $u$  and  $\varepsilon$  are mutually orthogonal (at all lags) white-noise processes. Calculating  $h(e^{-i\omega})$ , we have

$$\begin{aligned} h(e^{-i\omega}) &= 1 - e^{-i\omega} = e^{-i\omega/2}(e^{i\omega/2} - e^{-i\omega/2}) \\ &= e^{-i\omega/2} 2i \sin(\omega/2) = 2e^{-i\omega/2} e^{i\pi/2} \sin(\omega/2) \\ &= 2e^{i(\pi/2 - \omega/2)} \sin(\omega/2) \end{aligned}$$

For  $0 < \omega < \pi$ , the phase angle is positive, implying that the output  $y$  leads  $x$  at all frequencies between zero and  $\pi$ . In spite of the fact that  $y$  leads at all of these 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

$$y_t = \sum_{j=-\infty}^{\infty} h_j x_{t-j} + \varepsilon_t, \quad y_t = \lambda y_{t-1} + u_t$$

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

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

The cross spectrum between  $y$  and  $x$  is calculated to be

$$\begin{aligned} g_{yx}(e^{-i\omega}) &= \{h_0 + h_1(e^{i\omega} + e^{-i\omega}) + h_2(e^{2i\omega} + e^{-2i\omega}) + \dots\} g_x(e^{-i\omega}) \\ &= \left( h_0 + 2 \sum_{j=1}^{\infty} h_j \cos \omega j \right) g_x(e^{-i\omega}) \end{aligned}$$

which is *real* for all  $\omega$ . Therefore, the phase shift  $\theta(\omega) = 0$  for all  $\omega$ , so that  $y$  and  $x$  are perfectly in phase at all frequencies. Despite this, by using a theorem due to Sims (see pp. 277-287) 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 the  $h_j$  being two-sided and of Sims's theorem 2, which we will describe in detail presently.

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.

#### 8. ANALYSIS OF SOME FILTERS: THE SLUTSKY EFFECT AND KUZNETS' TRANSFORMATIONS

Relation (26) can be used to show the famous "Slutsky effect" (1937). Slutsky considered the effects of starting with a white noise  $\varepsilon_t$ , taking a two-period moving sum  $n$  times, and then taking first differences  $m$  times. That is, Slutsky considered forming the series

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

and

$$y_t = (1 - L)(1 - L) \cdots (1 - L)Z_t = (1 - L)^m Z_t = (1 + L)^n (1 - L)^m \varepsilon_t. \quad (44)$$

Applying (26) to (44); we have

$$\begin{aligned} g_y(e^{-i\omega}) &= (1 + e^{i\omega})^n (1 + e^{-i\omega})^n (1 - e^{i\omega})^m (1 - e^{-i\omega})^m \sigma_\varepsilon^2 \\ &= [(1 + e^{i\omega})(1 + e^{-i\omega})]^n [(1 - e^{i\omega})(1 - e^{-i\omega})]^m \sigma_\varepsilon^2 \\ &= [2 + (e^{i\omega} + e^{-i\omega})]^n [2 - (e^{i\omega} + e^{-i\omega})]^m \sigma_\varepsilon^2 \\ &= \sigma_\varepsilon^2 2^n [1 + \cos \omega]^n 2^m [1 - \cos \omega]^m. \end{aligned} \quad (45)$$

Consider first the special case where  $m = n$ . Then (45) becomes

$$g_y(e^{-i\omega}) = \sigma_\varepsilon^2 4^n [1 - \cos^2 \omega]^n = \sigma_\varepsilon^2 4^n [\sin^2 \omega]^n. \quad (46)$$

On  $(0, \pi)$ , the spectrum of  $y$  has a peak at  $\omega = \pi/2$  since there  $\sin \omega = 1$ . Notice that since  $\sin \omega \leq 1$ , (45) implies that as  $n$  becomes large, the peak in the spectrum of  $y$  at  $\pi/2$  becomes sharp. In the limit, as  $n \rightarrow \infty$ , 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 (45) and set  $dg_y(e^{-i\omega})/d\omega$  equal to zero in order to locate the peak in the spectrum:

$$\begin{aligned} dg_y/d\omega &= \sigma_\varepsilon^2 2^{m+n} \{n[1 - \cos \omega]^m [1 + \cos \omega]^{n-1} (-\sin \omega) \\ &\quad + m(1 - \cos \omega)^{m-1} (\sin \omega) [1 + \cos \omega]^n\} \\ &= \sigma_\varepsilon^2 2^{m+n} \sin \omega \{(1 - \cos \omega)^{m-1} (1 + \cos \omega)^{n-1} \\ &\quad \cdot [m(1 + \cos \omega) - n(1 - \cos \omega)]\}. \end{aligned}$$

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

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

which implies

$$\cos \omega = \frac{1 - (m/n)}{1 + (m/n)}, \quad \text{or} \quad \omega = \cos^{-1} \left( \frac{1 - (m/n)}{1 + (m/n)} \right)$$

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  $n \rightarrow \infty$ . This means that as  $n \rightarrow \infty$ ,  $y$  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  $\varepsilon_t$ , a series with "cycles" is obtained.

Another use of (26) is in the analysis of transformations that have been applied to data. An example is Howrey's (1968) 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

$$g_y(e^{-i\omega}) = A(e^{-i\omega})A(e^{i\omega})B(e^{-i\omega})B(e^{i\omega})g_x(e^{-i\omega}). \quad (47)$$

We have

$$A(e^{-i\omega}) = \frac{1}{5} \sum_{j=-2}^2 e^{-i\omega j} = \frac{1}{5} \frac{e^{i\omega 2} - e^{-i\omega 3}}{1 - e^{-i\omega}}.$$

Thus,

$$\begin{aligned} A(e^{-i\omega})A(e^{i\omega}) &= \frac{(\frac{1}{5})^2 (e^{i\omega 2} - e^{-i\omega 3})(e^{-i\omega 2} - e^{i\omega 3})}{(1 - e^{-i\omega})(1 - e^{i\omega})} \\ &= \frac{(\frac{1}{5})^2 (2 - (e^{i\omega 5} + e^{-i\omega 5}))}{2 - (e^{i\omega} + e^{-i\omega})} \\ &= \frac{(\frac{1}{5})^2 2(1 - \cos 5\omega)}{2(1 - \cos \omega)} = \frac{(\frac{1}{5})^2 (1 - \cos 5\omega)}{1 - \cos \omega}. \end{aligned}$$

Next, we have  $B(e^{-i\omega}) = (e^{+i\omega 5} - e^{-i\omega 5})$ , so that

$$\begin{aligned} B(e^{i\omega})B(e^{-i\omega}) &= (e^{i\omega 5} - e^{-i\omega 5})(e^{-i\omega 5} - e^{i\omega 5}) \\ &= (2 - (e^{i\omega 10} + e^{-i\omega 10})) = 2(1 - \cos 10\omega). \end{aligned}$$

So it follows from (26) that

$$g_y(e^{-i\omega}) = \frac{(\frac{1}{5})^2 (1 - \cos 5\omega) 2}{(1 - \cos \omega)} (1 - \cos 10\omega) g_x(e^{-i\omega}) = G(\omega) g_x(e^{-i\omega}).$$

where  $G(\omega) = 2[(\frac{1}{5})^2 (1 - \cos 5\omega)(1 - \cos 10\omega)/(1 - \cos \omega)]$ . The term  $G(\omega)$  is graphed in Figure 4. It has zeros at values where  $\cos 5\omega = 1$  and where  $\cos 10\omega = 1$ . The first condition occurs on  $[0, \pi]$  where

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

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

$$10\omega = 0, 2\pi, 4\pi, 6\pi, 8\pi, 10\pi \quad \text{or} \quad \omega = 0, \frac{1}{5}\pi, \frac{2}{5}\pi, \frac{3}{5}\pi, \frac{4}{5}\pi, \pi.$$

So  $G(\omega)$  has zeros at  $\omega = 0, \frac{1}{5}\pi, \frac{2}{5}\pi, \frac{3}{5}\pi, \frac{4}{5}\pi$ , and  $\pi$ .

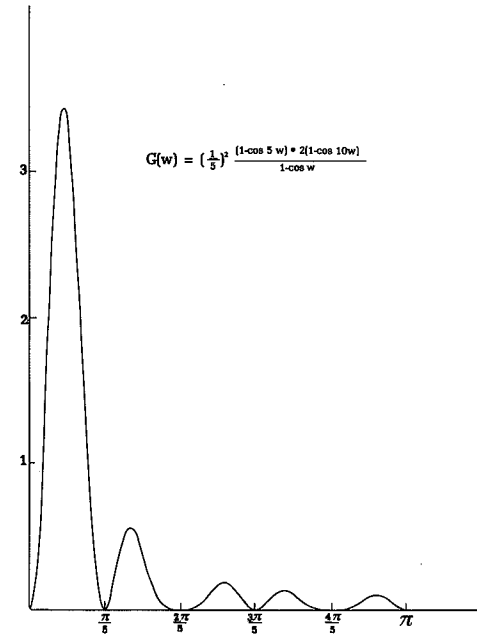


FIGURE 4

From the graph of  $G(\omega)$ , 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; i.e., 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 4 corresponds to a cycle of about  $20\frac{1}{4}$  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.

#### 9. A SMALL KIT OF $h(e^{-i\omega})$ 's

In order to provide some feel for the effects of various commonly used filters Figure 5 reports the amplitude and phase of  $h(e^{-i\omega})$  for various  $h(L)$  lag distributions.

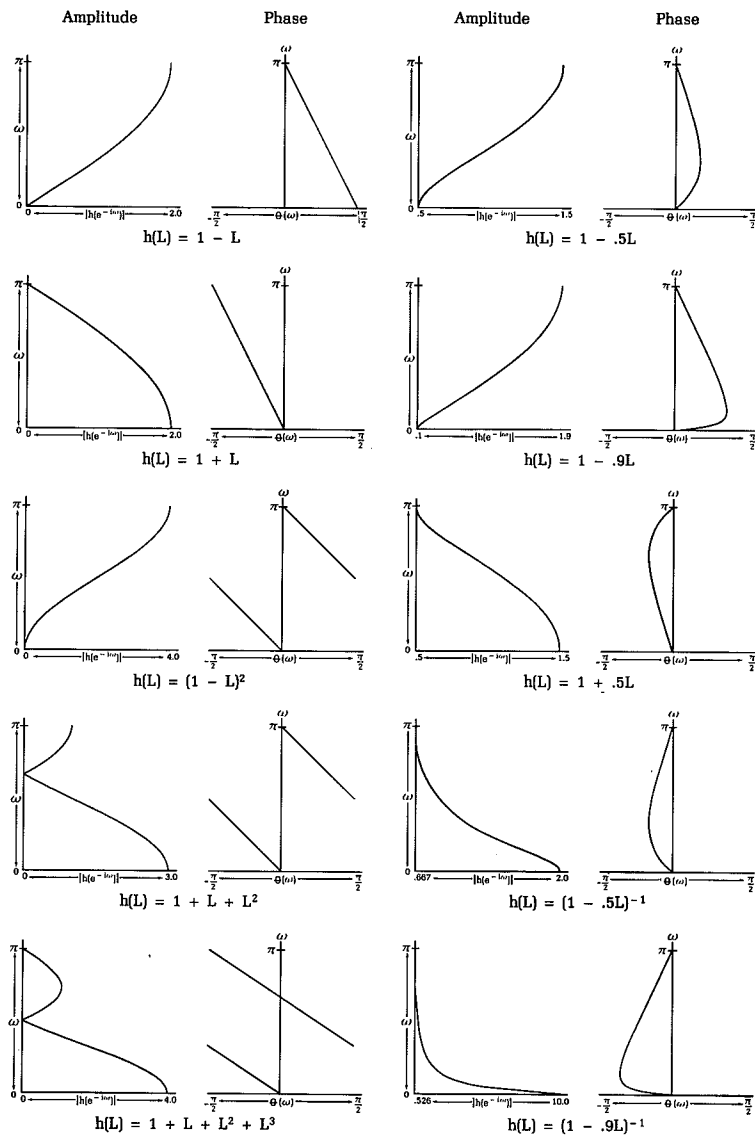


FIGURE 5 Frequency response functions.

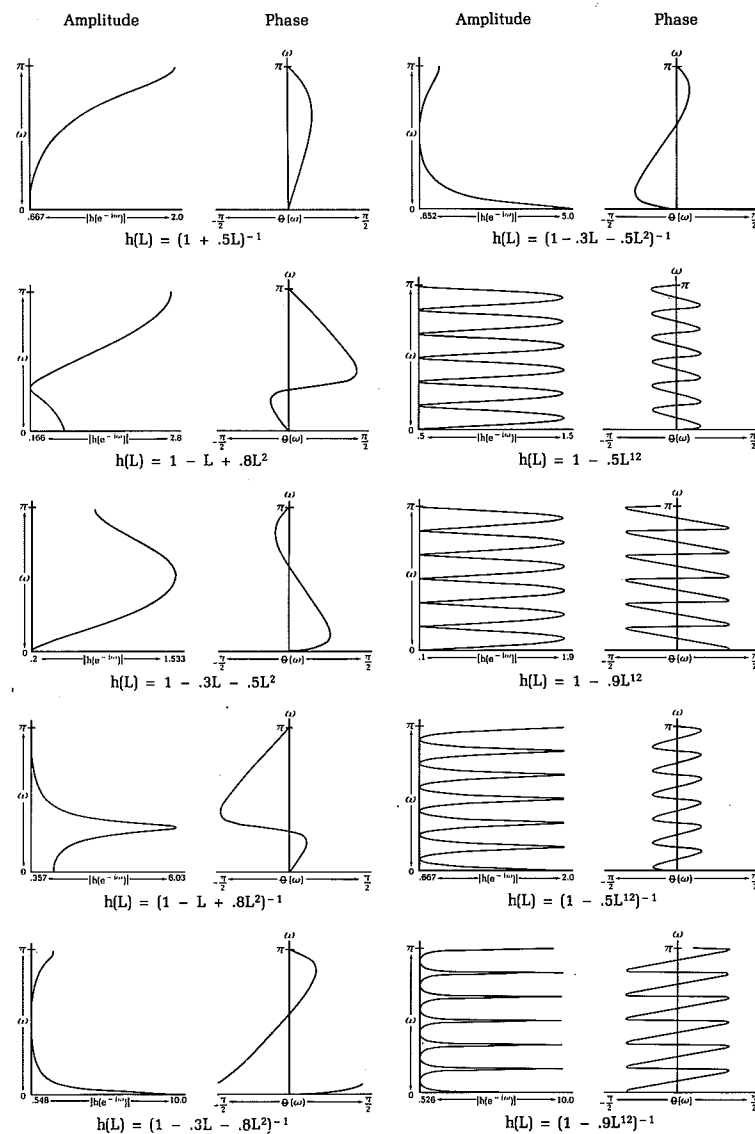


FIGURE 5 (Continued)

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

$$h(e^{-i\omega}) = 2e^{i(\pi/2 - \omega/2)} \sin(\omega/2),$$

as the graphs confirm.

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

$$h(e^{-i\omega}) = 1 + e^{-i\omega} = e^{-i\omega/2}(e^{+i\omega/2} + e^{-i\omega/2}) = 2e^{-i\omega/2} \cos(\omega/2),$$

which again agrees with our graphs.

Notice that for  $h(L) = (1 - t_1 L - t_2 L^2)^{-1}$ , we have chosen  $(t_1, t_2)$  in the regions of peaked spectra of our Figure 3. Notice that as required,  $h(e^{-i\omega})$  is characterized by peaks. (See Figure 3.)

#### 10. 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  $(\theta$  in the polar form of the root  $re^{i\theta}$ ) equal to the frequency in question. A single series is said to contain a *business cycle* 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 (1966) the "typical spectral shape" of an economic variable and is illustrated by the logarithms of the spectral densities of real GNP, the unemployment rate, the real wage, the Baa rate, and output per man-hour in Figure 1. The generally downward sweeping spectrum is characteristic of a covariogram that is dominated by high, positive, low-order serial correlation. Notice that the inflation rate and change in the real money supply do not display the typical spectral shape, a characteristic that might have been anticipated from our study of the effects of applying the first difference filter  $1 - L$ . All of the series except the

unemployment rate, which has been seasonally adjusted, display spectral peaks in the vicinity of four and two quarters, which is symptomatic of a seasonal pattern of serial correlation. "Seasonal adjustment" is a process of operating on a series with a filter  $h(L)$  that is designed to diminish the seasonal frequencies near four and two quarters, while leaving the remaining frequencies as unaffected as possible. Notice how this procedure has "succeeded" for the unemployment rate and produced dips in the spectrum near four and two quarters.

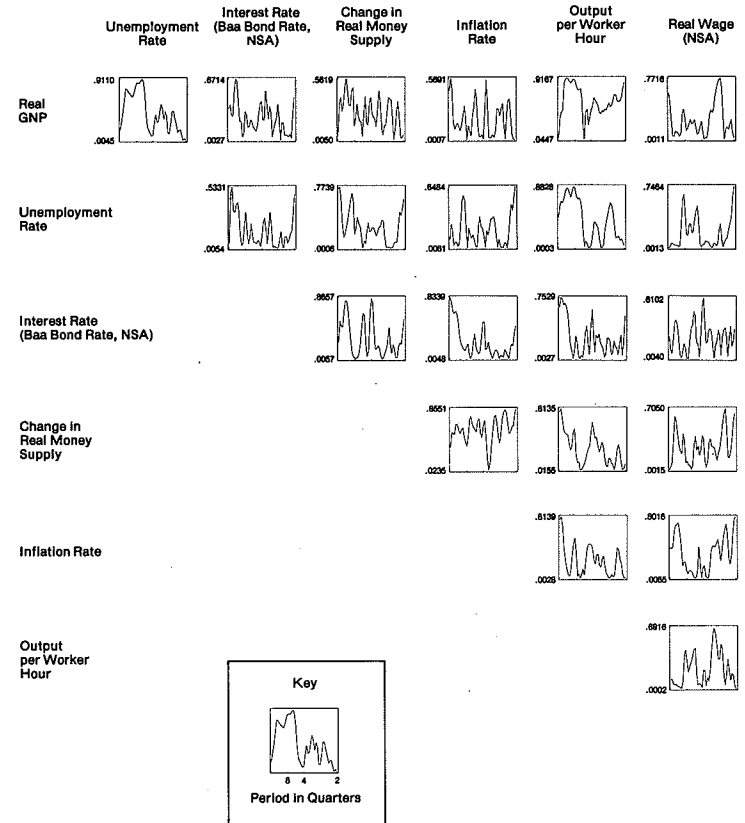


FIGURE 6 Coherences

Notice how real GNP has no spectral peak in the business cycle range, while output per man-hour and the unemployment rate have only very modest peaks, this despite the fact that the sample paths of all three reflect "the business cycle." As mentioned earlier, the fact that a spectrum does not display a peak at the business cycle frequencies should not be taken to mean that the series did not experience any fluctuations associated with the business cycle. On the contrary, as Figure 2a 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 interval  $(0, \pi)$ . 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 tentative 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. Figure 6 reports estimated coherences for the six variables graphed in Figure 1 over the period 1948I–1976IV. Notice the high pairwise coherences among the unemployment rate, real GNP, and output per man-hour at the low business cycle frequencies.

### 11. REPRESENTATION THEORY

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

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

In this section we reverse this procedure and start 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,  $\{x_{t-1}, x_{t-2}, \dots, x_{t-n}\}$ :

$$\hat{x}_t^n = \sum_{i=1}^n a_i^n x_{t-i} = P[x_t | x_{t-1}, \dots, x_{t-n}] \quad \text{or} \quad x_t = \hat{x}_t^n + \varepsilon_t^n$$

where  $E\varepsilon_t^n x_{t-i} = 0$  for  $i = 1, \dots, n$  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  $a_i^n$  from the least squares normal equations.<sup>11</sup>

As  $n$  is increased toward infinity, it is possible to show that the sequence of projections  $\{\hat{x}_t^n\}$  converges to a random variable  $\hat{x}_t$  in the "mean square" sense that<sup>12</sup>

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

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 on 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.<sup>13</sup> We write the projection of  $x_t$  on the space spanned by the infinite set  $(x_{t-1}, x_{t-2}, \dots)$  as

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

and have the decomposition of  $x_t$  as

$$x_t = P[x_t | x_{t-1}, x_{t-2}, \dots] + \varepsilon_t \quad (48)$$

where  $\varepsilon_t$  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  $\varepsilon_t$  is the limit as  $n \rightarrow \infty$  of  $\varepsilon_t^n$ , i.e.,  $\lim_{n \rightarrow \infty} E(\varepsilon_t - \varepsilon_t^n)^2 = 0$ .

We can now state an important decomposition theorem due to Wold.<sup>14</sup>

**Theorem:** Let  $\{x_t\}$  be any covariance stationary stochastic process with  $E x_t = 0$ . Then it can be written as

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

where  $d_0 = 1$  and 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

<sup>11</sup> The  $a_i^n$  will be unique only if there are no linear dependencies across the  $x_{t-i}$ . The projection of  $x_t$  on the space spanned by  $\{x_{t-1}, \dots, x_{t-n}\}$  is unique even without that condition.

<sup>12</sup> It is not necessarily true that the sequence of  $a_i^n$  settles down nicely as  $n \rightarrow \infty$ , only that successive  $\hat{x}_t^n$  get closer to each other and to  $\hat{x}_t$  as  $n \rightarrow \infty$ .

<sup>13</sup> For a proof, see Anderson (1971, p. 419).

<sup>14</sup> See Wold (1938). The proof given here parallels that given by Anderson (1971). The reader familiar with Hilbert spaces is urged to read Anderson at this point.

$\{\varepsilon\}$  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.,  $\eta_t$  is linearly deterministic. Furthermore,  $\varepsilon_t = x_t - P[x_t | x_{t-1}, x_{t-2}, \dots]$ . *Proof:* We let  $\varepsilon_t$  be the same  $\varepsilon_t$  as appears in (48), so that

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

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

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

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  $(\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-m})$  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  $\varepsilon_{t-j}$  are mutually orthogonal, the  $d_j$  are given by

$$d_j = (Ex_t \varepsilon_{t-j}) / \sigma^2, \quad \sigma^2 = E\varepsilon_t^2.$$

Notice that since  $\varepsilon_t = x_t - P[x_t | x_{t-1}, x_{t-2}, \dots]$  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  $\varepsilon$ 's are orthogonal, the  $d_j$  do not depend on  $m$ . Now calculate the variance of the prediction error, which is

$$\begin{aligned} E\left(x_t - \sum_{j=0}^m d_j \varepsilon_{t-j}\right)^2 &= Ex_t^2 - 2 \sum_{j=0}^m d_j Ex_t \varepsilon_{t-j} + E\left(\sum_{j=0}^m d_j^2 \varepsilon_{t-j}^2\right) \\ &= Ex_t^2 - 2\sigma^2 \sum_{j=0}^m \left(\frac{Ex_t \varepsilon_{t-j}}{\sigma^2}\right) + \sigma^2 \sum_{j=0}^m \left(\frac{Ex_t \varepsilon_{t-j}}{\sigma^2}\right)^2 \\ &= Ex_t^2 - \sigma^2 \sum_{j=0}^m d_j^2 \geq 0, \end{aligned}$$

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.<sup>15</sup>

Now define the process  $\eta_t$  by

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

Notice that for  $s \leq t$  we have

$$\begin{aligned} E\eta_t \varepsilon_s &= Ex_t \varepsilon_s - E \sum_{j=0}^{\infty} d_j \varepsilon_s \varepsilon_{t-j} = Ex_t \varepsilon_s - d_{t-s} E\varepsilon_s^2 \\ &= Ex_t \varepsilon_s - Ex_t \varepsilon_s = 0. \end{aligned}$$

In addition  $E\eta_t \varepsilon_s = 0$  for all  $s > t$  because  $\varepsilon_s$  is orthogonal to all  $x$ 's dated earlier than  $s$  and by construction  $\eta_t$  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  $\eta_t$  is orthogonal to  $\varepsilon_t$ ,  $\eta_t$  must lie in the space spanned by  $\{x_{t-1}, x_{t-2}, \dots\}$  since square summable<sup>16</sup> linear combinations of  $\{x_{t-1}, x_{t-2}, \dots\}$  form the space of all random variables orthogonal to  $\varepsilon_t$ .<sup>17</sup> This implies that  $\eta_t$  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  $\{x_{t-1}, x_{t-2}, \dots\}$  to get

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

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

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

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

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

<sup>15</sup> That is, the sequence of  $\sum_{j=0}^m d_j \varepsilon_{t-j}$  is a Cauchy sequence. In particular, for  $n > m$ ,

$$\begin{aligned} E\left(\sum_{j=0}^m d_j \varepsilon_{t-j} - \sum_{j=0}^n d_j \varepsilon_{t-j}\right)^2 &= E\left(\sum_{j=m+1}^n d_j^2 \varepsilon_{t-j}^2\right) \\ &= \sigma^2 \sum_{j=m+1}^n d_j^2. \end{aligned}$$

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.

<sup>16</sup> 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.

<sup>17</sup> This is an implication of the orthogonality principle. See Anderson (1971).

Subtracting this from the definition of  $\eta_t$  gives

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

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  $\eta_t$  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  $\eta_t$  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  $\varepsilon_t$  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

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

where  $\varepsilon_t$  is a covariance stationary, serially uncorrelated process with mean zero and variance  $\sigma_\varepsilon^2$ ;  $\sum_{j=0}^{\infty} d_j^2 < \infty$ ;  $a_i$  and  $b_i$  are random variables orthogonal to the entire  $\varepsilon$  process and satisfying  $Ea_i = Eb_i = Ea_i b_j = 0$  for all  $i, j$ ,  $Ea_i a_j = Eb_i b_j = 0$  for all  $i \neq j$ , and  $Ea_i^2 = Eb_i^2 = \sigma_i^2$ ; the  $\lambda_i$  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  $\sum_{j=0}^{\infty} d_j \varepsilon_{t-j}$  at all lags, and is easily deduced<sup>18</sup> to have covariogram given by  $\sum_{i=1}^n \sigma_i^2 \cos \lambda_i \tau$ . As we have seen, the covariogram of

<sup>18</sup> For example, let  $x(t) = a \cos \lambda_1 t + b \sin \lambda_1 t$  where  $Ea = Eab = Eb = 0$ ,  $Ea^2 = Eb^2 = \sigma^2$ . Then

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

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

$$Ex(t_1)x(t_2) = \sigma^2 \cos \lambda(t_1 - t_2) \quad \text{or} \quad Ex(t)x(t - T) = \sigma^2 \cos \lambda T.$$

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

$\sum_{j=0}^{\infty} d_j \varepsilon_{t-j}$  has generating function  $\sigma_\varepsilon^2 d(z)d(z^{-1})$ . The spectral density of the deterministic part turns out to be not 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

$$\begin{aligned} \sigma^2 \sum_{\tau=-\infty}^{\infty} \cos \lambda \tau e^{-i\omega \tau} &= \sigma^2 \sum_{\tau=-\infty}^{\infty} \left( \frac{e^{i\lambda \tau} + e^{-i\lambda \tau}}{2} \right) e^{-i\omega \tau} \\ &= \sigma^2 \sum_{\tau=-\infty}^{\infty} \left( \frac{e^{i(\lambda - \omega)\tau} + e^{-i(\lambda + \omega)\tau}}{2} \right). \end{aligned}$$

Notice that the first term can be written

$$\sum_{\tau=-\infty}^{\infty} e^{i(\lambda - \omega)\tau} = 1 + \sum_{\tau=1}^{\infty} (e^{i(\lambda - \omega)\tau} + e^{-i(\lambda - \omega)\tau}) = 1 + 2 \sum_{\tau=1}^{\infty} \cos(\lambda - \omega)\tau.$$

The series  $\sum_{\tau=1}^{\infty} \cos(\lambda - \omega)\tau$  is not a convergent series, so 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  $\delta(\omega)$  be the delta generalized function which has "infinite height and unit mass" at  $\omega = 0$  and is zero everywhere else. That is,  $\delta(\omega)$  is defined by

$$\int_{-\infty}^{\infty} \delta(\omega) g(\omega) d\omega = g(0),$$

which must hold for all "test functions"  $g(\omega)$  that are continuous at  $\omega = 0$ . Then the spectral density of a process with covariogram  $\sigma^2 \cos \lambda \tau$  is defined as

$$f(\omega) = 2\pi \left( \frac{1}{2} \sigma^2 \delta(\omega - \lambda) + \frac{1}{2} \sigma^2 \delta(\omega + \lambda) \right).$$

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

$$\begin{aligned} c(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\omega) e^{i\omega \tau} d\omega = \frac{\sigma^2}{2} \left( \int_{-\infty}^{\infty} \delta(\omega - \lambda) e^{i\omega \tau} d\omega + \int_{-\infty}^{\infty} \delta(\omega + \lambda) e^{i\omega \tau} d\omega \right) \\ &= \sigma^2 \left( \frac{e^{i\lambda \tau} + e^{-i\lambda \tau}}{2} \right) = \sigma^2 \cos \lambda \tau. \end{aligned}$$

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

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



so the spectral density function of the deterministic part is zero except for the singular points  $\omega = \pm \lambda_i$ ,  $i = 1, \dots, n$ , at which the spectrum has mass  $\sigma_i^2/2$ . The spectral density thus has "spikes" at the points  $\omega = \pm \lambda_i$ .<sup>19</sup>

## 12. LINEAR LEAST SQUARES PREDICTION<sup>20</sup>

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  $\eta_t$  has been removed.<sup>21</sup> Wold's theorem says that any indeterministic covariance stationary stochastic process  $x_t$  has the *moving average representation*

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

or

$$x_t = d(L)\varepsilon_t, \quad d(L) = \sum_{j=0}^{\infty} d_j L^j \quad (49)$$

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  $\{x_{t-1}, x_{t-2}, \dots\}$ , i.e.,  $\varepsilon_t = x_t - P[x_t | x_{t-1}, x_{t-2}, \dots]$ . (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  $d(L) = \sum_{j=0}^{\infty} d_j L^j$ , a necessary and sufficient condition for  $d(L)$  to have such a one-sided inverse is that the roots  $\mu$  of  $\sum_{j=0}^{\infty} d_j \mu^j = 0$  all lie outside the unit circle, i.e., all have absolute values exceeding 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 (49) with  $a(L) = d(L)^{-1}$  gives

$$a(L)x_t = \varepsilon_t, \quad a(L) = a_0 - \sum_{j=1}^{\infty} a_j L^j \quad (50)$$

or

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

Since  $d_0$  is unity, it turns out that  $a_0$  is unity also. Equation (50) is termed the autoregressive representation for  $x_t$ . While every linearly indeterministic co-

<sup>19</sup> 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  $\omega$ 's of positive length, or is "too close" to zero over such an interval. Heuristically, this second possible way of being deterministic is suggested by the Kolmogorov formula for the one-step-ahead prediction error variance  $\sigma_t^2 = \exp[(2\pi)^{-1} \int_{-\pi}^{\pi} \ln g(e^{-i\omega}) d\omega]$  where  $g(e^{-i\omega})$  is the spectral density. See Whittle (1963, p. 26).

<sup>20</sup> A key reference on the subject of this section is Whittle (1963).

<sup>21</sup> For example, by suitable detrending and seasonal adjustment.

variance 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 are dealing with a member of this class.<sup>22</sup>

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.,

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

Now project both sides of (49) against  $\{x_{t-1}, x_{t-2}, \dots\}$  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  $\varepsilon_t$  is orthogonal to lagged  $x$ 's; and since  $P_{t-1}\varepsilon_{t-j} = \varepsilon_{t-j}$  for all  $j \geq 1$ , because  $\varepsilon_{t-j}$  is in the space spanned by  $\{x_{t-1}, x_{t-2}, \dots\}$ . We write the above equation as

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

where  $(\ )_+$  means "ignore negative powers of  $L$ ," i.e.,  $(\sum_{j=-\infty}^{\infty} h_j L^j)_+ \equiv \sum_{j=0}^{\infty} h_j L^j$ . 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

$$P_{t-1}x_t = \left(\frac{d(L)}{L}\right)_+ \frac{1}{d(L)} x_{t-1}, \quad (51)$$

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 (49) against  $\{x_{t-k}, x_{t-k-1}, \dots\}$  to get

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

$$P_{t-k}x_t = \left(\frac{d(L)}{L^k}\right)_+ \frac{1}{d(L)} x_{t-k} \quad (52)$$

which generalizes formula (51). Equation (52) is the Wiener-Kolmogorov formula for  $k$ -step-ahead linear least squares predictions.

<sup>22</sup> We remarked earlier that in general the sequence of the  $a_j^n$  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  $n \rightarrow \infty$ . However, under the roots condition given in the text, the  $a_j^n$  do converge. In particular, they converge to the  $a_j$  of Equation (50), so that  $\lim_{n \rightarrow \infty} a_j^n = a_j$  for all  $j = 1, 2, \dots$ .

A. Some Examples<sup>23</sup>

**First-Order Markov:** Consider the first-order autoregressive process  $(1 - \lambda L)x_t = \varepsilon_t$ ,  $\varepsilon_t$  white noise,  $|\lambda| < 1$ ,  $\varepsilon_t = x_t - P[x_t | x_{t-1}, \dots]$ ; we can write  $x_t = (1/(1 - \lambda L))\varepsilon_t$ . We have

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

More generally,

$$P_{t-k}x_t = [L^{-k}(1 + \lambda L + \dots)]_+(1 - \lambda L)x_{t-k} = \lambda^k x_{t-k}.$$

Thus we have

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

**First-Order Moving Average:** Suppose  $x_t = (1 + \beta L)\varepsilon_t$ ,  $\varepsilon_t$  white,  $|\beta| < 1$ ,  $\varepsilon_t = x_t - P[x_t | x_{t-1}, \dots]$ . Then we have

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

We also have that for  $k \geq 2$ ,

$$P_{t-k}x_t = [L^{-k}(1 + \beta L)]_+\left(\frac{1}{1 + \beta L}\right)x_{t-1} = 0,$$

which can also be seen directly by projecting on  $\{x_{t-k}, x_{t-k-1}, \dots\}$  both sides of  $x_t = (1 + \beta L)\varepsilon_t$ .

**First-Order Moving Average, Autoregressive:** Suppose we have

$$\begin{aligned} x_t &= \left(\frac{1 + aL}{1 - \beta L}\right)\varepsilon_t, \quad \varepsilon_t \text{ white}, \quad |a| < 1, \quad |\beta| < 1, \\ \varepsilon_t &= x_t - P[x_t | x_{t-1}, \dots]. \end{aligned}$$

We then have

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

<sup>23</sup> In these examples we continue to assume that  $Ex_t = E\varepsilon_t = 0$ . Modifying the formulas to account for a nonzero mean of  $x_t$  is trivial and involves adding constant terms to the formulas.

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  $\beta \rightarrow 1$  (from below, in order to assure that the roots condition  $|\beta| < 1$  is met),  $P_{t-1}x_t$  approaches

$$P_{t-1}x_t = \{(1 + a)/(1 + aL)\}x_{t-1},$$

which with  $a < 0$  is equivalent with Cagan (1956) adaptive expectations scheme

$$P_{t-1}x_t = \{(1 - \lambda)/(1 - \lambda L)\}x_{t-1},$$

with  $a = -\lambda$ . Notice that as  $\beta \rightarrow 1$  (from below), we approach the situation in which  $(1 - L)x_t = (1 + aL)\varepsilon_t$ , 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

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

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

$$P_t x_{t+k} = ((1 + a)/(1 + aL))x_t,$$

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 Muth (1960), 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 Friedman had earlier used to estimate permanent income in empirical work.

## 13. 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

$$x_t = d_0 \varepsilon_t + d_1 \varepsilon_{t-1} \quad (53)$$

where  $d_0$  and  $d_1$  are to be determined, and  $\varepsilon_t$  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 (53) with  $\{\varepsilon_t\}$  being a white noise with variance  $\sigma_\varepsilon^2$ , it is straightforward to calculate

$$c(0) = (d_0^2 + d_1^2)\sigma_\varepsilon^2, \quad c(1) = (d_0 d_1)\sigma_\varepsilon^2. \quad (54)$$

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  $\sigma_\varepsilon^2$ . The equations are graphed for fixed  $\sigma_\varepsilon^2$  and  $c(1) > 0$  in Figure 7. In general, the two equations determine two pairs of solutions, one pair consisting of  $d_0 = \alpha > \beta = d_1$  and  $d_1 = \alpha > \beta = d_0$ , where  $\alpha$  and  $\beta$  are the positive scalars depicted in Figure 7; the second pair is the reflection of the first pair in the negative quadrant. As  $\sigma_\varepsilon^2$  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  $d_0 = 1$ . (We are free to normalize by choosing  $\sigma_\varepsilon^2$  so that  $d_0 = 1$ .) Which of the two solutions with  $d_0 > 0$  should be chosen? The answer comes

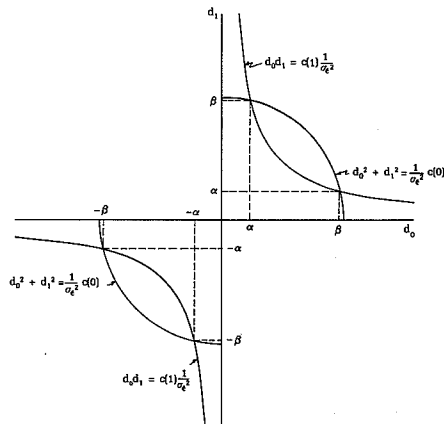


FIGURE 7

from the condition that the derived  $\varepsilon_t$  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

$$x_t = d_0 \varepsilon_t + d_1 \varepsilon_{t-1} \quad \text{or} \quad \varepsilon_t = (1/d_0)x_t - (d_1/d_0)\varepsilon_{t-1}, \quad (55)$$

so that  $\varepsilon_t$  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  $|-d_1/d_0| > 1$ . The forward solution of the difference equation (54) is "stable" if  $d_1 > d_0$ . That is, as we saw earlier, we can write

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

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  $\varepsilon_t$  as a convergent (square summable) series of future  $x$ 's. Thus, if  $d_1 > d_0$ , the associated  $\varepsilon_t$  does not lie in the space spanned by current and lagged  $x$ 's. However, if  $d_0 > d_1$ , the associated  $\varepsilon_t$  process does lie in the space spanned by current and lagged  $x$ 's,<sup>24</sup> which is the condition that will always result in choosing the correct roots of (54). The general principle is this: in selecting among the sequences  $\{d_0, d_1, d_2, \dots\}$  that solve the equations that are the general counterparts of (54), choose the representation in which  $d_0 \sigma_\varepsilon^2$  is maximal. 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  $\varepsilon_t$  with this property are said to be the fundamental white-noise process for  $x_t$ . Ordinarily, we normalize by choosing  $\sigma_\varepsilon^2$  so that  $d_0 = 1$ . In this case  $\varepsilon_t$  equals the one-step-ahead prediction error for  $x_t$ .

As a practical matter, solving equations of the form (54) 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  $a_i^n$  in

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

<sup>24</sup> By an appropriate limiting argument it can be shown that  $\varepsilon_t$  lies in that space even if  $d_0 = d_1$ .

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

$$d^n(L) = \left(1 - \sum_{i=1}^n a_i^n L^i\right)^{-1}.$$

By making  $n$  large enough, an arbitrarily good approximation<sup>25</sup> to  $d(L)$  can be obtained.

#### 14. 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.<sup>26</sup>

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

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

It follows that

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

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

$$P_t x_{t+k+1} = h_0 P_t x_{t+k} + h_1 P_t x_{t+k-1} + \dots + h_{k-1} P_t x_{t+1} + \sum_{i=0}^{\infty} h_{k+i} x_{t-i}. \quad (56)$$

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 of the formula for the one-step-ahead projection alone.

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

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

From application of (56) it follows that  $P_t x_{t+j} = \lambda^j x_t, j \geq 1$ .

#### 15. SOME APPLICATIONS TO RATIONAL EXPECTATIONS MODELS

Let us return to the example of Cagan's portfolio balance schedule, only now we assume that  $m_t$  is a covariance stationary stochastic process and the log of 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

$$m_t - p_t = \alpha P_t p_{t+1} - \alpha p_t, \quad \alpha < 0 \quad (57)$$

<sup>25</sup> Arbitrarily good in the sense that the variance of the  $\{\epsilon_t^n\}$  process can be made as close as desired to the variance of  $\{\epsilon_t\}$  by making  $n$  large enough.

<sup>26</sup> Interesting applications of the chain rule of forecasting occur in Shiller (1972).

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  $B P_{t-1} x_{t+j} \equiv P_{t-1} x_{t+j-1}$  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 with  $B$  gives

$$\left(1 - \frac{\alpha - 1}{\alpha} B\right) P_{t-1} p_t = \frac{1}{\alpha} P_{t-1} m_{t-1}.$$

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

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

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

It is natural to guess and can be verified directly that a solution to the stochastic difference equation (57) is then<sup>27</sup>

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

Now suppose that  $m_t$  has the moving average representation

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

where  $\sum_{j=0}^{\infty} d_j^2 < \infty$  and  $\epsilon_t$  is fundamental for  $m$ . Then we have, applying (52),

$$\begin{aligned} p_t &= \frac{1}{1 - \alpha} \sum_{j=0}^{\infty} \left( \frac{\alpha}{\alpha - 1} \right)^j \left[ \frac{d(L)}{L^j} \right]_+ \epsilon_t \\ &= \frac{1}{1 - \alpha} \left[ d(L) + \frac{d(L)}{L} \left( \frac{\alpha}{1 - \alpha} \right) + \frac{d(L)}{L^2} \left( \frac{\alpha}{1 - \alpha} \right)^2 + \dots \right]_+ \epsilon_t \\ &= \frac{1}{1 - \alpha} \left[ \sum_{j=0}^{\infty} \left( \frac{\alpha}{1 - \alpha} \right)^j \frac{d(L)}{L^j} \right]_+ \frac{1}{d(L)} m_t, \end{aligned}$$

<sup>27</sup> In this and the following example we set "transient terms" of the form  $c\lambda^t$  to zero because we are interested in obtaining solutions that are covariance stationary processes.

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

Let us now consider the supply-demand example of Chapter IX where  $x_t$  is now a covariance stationary, indeterministic random process with mean zero and moving average representation  $x_t = d(L)e_t$ . Our system is naturally modified to become

$$\begin{aligned} C_t &= -\beta p_t, & \beta > 0 \\ Y_t &= \gamma P_{t-1} p_t + x_t, & \gamma > 0 \\ I_t &= \alpha(P_t p_{t+1} - p_t), & \alpha > 0 \\ Y_t &= C_t + I_t - I_{t-1}, \end{aligned}$$

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

$$(\gamma + \alpha)P_{t-1}p_t + (\alpha + \beta)p_t = \alpha P_t p_{t+1} + \alpha p_{t-1} - x_t. \quad (58)$$

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

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

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  $\phi = ((\beta + \gamma)/\alpha) + 2 > 0$ . Multiplying by  $B$  gives

$$\begin{aligned} (1 - \phi B + B^2)P_{t-1}p_t &= \alpha^{-1}P_{t-1}x_{t-1} \\ (1 - \lambda^{-1}B)(1 - \lambda B)P_{t-1}p_t &= \alpha^{-1}P_{t-1}x_{t-1} \end{aligned} \quad (59)$$

where  $|\lambda| < 1$  satisfies  $\lambda + \lambda^{-1} = \phi$ . To ensure covariance stationarity of the solution, we shall insist that all lag distributions be square summable. Operating on both sides of (59) with the forward inverse of  $(1 - \lambda^{-1}B)$  gives

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

or

$$P_{t-1}p_t - \lambda p_{t-1} = \frac{-\lambda}{\alpha} \sum_{i=0}^{\infty} \lambda^i P_{t-1}x_{t+i}.$$

This solution for  $P_{t-1}p_t$  suggests that the solution for  $p_t$  that makes  $p_t$  depend only on current and lagged information is

$$p_t = \lambda p_{t-1} - \frac{\lambda}{\alpha} \sum_{i=0}^{\infty} \lambda^i P_t x_{t+i}. \quad (60)$$

That (60) is a solution can be verified by direct substitution into (58). We can eliminate  $P_t x_{t+i}$  by using the Wiener-Kolmogorov formula to get

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

This is the solution to the stochastic difference equation (58) which expresses  $p_t$  as a function of current and lagged  $x$ 's and  $p$ 's, and which gives a covariance stationary process for  $p_t$ .

## 16. VECTOR STOCHASTIC DIFFERENCE EQUATIONS

Let  $x_t$  be an  $(n \times 1)$ -vector wide-sense stationary stochastic process that is governed by the matrix difference equation

$$C(L)x_t = \varepsilon_t \quad (61)$$

where  $\varepsilon_t$  is now an  $n \times 1$  vector of white noises with means of zero and contemporaneous covariance matrix  $E\varepsilon_t \varepsilon_t' = V$ , an  $n \times n$  matrix. We assume  $E\varepsilon_t \varepsilon_{t-s}' = 0_{n \times n}$  for all  $s \neq 0$ . In (61),  $C(L)$  is an  $n \times n$  matrix of (finite order) polynomials in the lag operator  $L$ :

$$C(L) = \begin{bmatrix} C_{11}(L) & C_{12}(L) & \cdots & C_{1n}(L) \\ \vdots & & & \\ C_{n1}(L) & \cdots & & C_{nn}(L) \end{bmatrix},$$

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 that satisfies

$$C(L)^{-1}C(L) = I_{n \times n}$$

where  $I_{n \times n}$  is the  $n \times n$  identity matrix. If it exists,  $C(L)^{-1}$  can be found as follows. Evaluate the matrix  $z$  transform  $C(z)$  at  $z = e^{-i\omega}$  to get  $C(e^{-i\omega})$ . Then invert  $C(e^{-i\omega})$ , frequency by frequency, to get  $C(e^{-i\omega})^{-1}$ . Finally, the matrix coefficients  $C(L)^{-1} = B(L) = \sum_{j=0}^{\infty} B_j L^j$ ,  $B_j$  being an  $n \times n$  matrix, can be found from the inversion formula

$$B_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} C(e^{-i\omega})^{-1} e^{i\omega j} d\omega,$$

where by integrating a matrix we mean to denote element-by-element integration.

The solution of (61) is found by premultiplying (61) by  $B(L)$  to obtain

$$x_t = B(L)\varepsilon_t. \quad (62)$$

The vector stochastic difference equation  $C(L)x_t = \varepsilon_t$  is said to be an *autoregressive representation* for the vector process  $x_t$ . The solution  $x_t = B(L)\varepsilon_t$  is said to be a *vector moving average representation* for the process  $x_t$ . The cross-spectral density matrix of the  $n \times 1$   $x_t$  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

$$g_{xx}(e^{-i\omega}) = B(e^{-i\omega})VB(e^{i\omega})' \quad (63)$$

where the prime denotes transposition. Formula (63) is analogous to the univariate equation (5), and can be derived by comparable methods.

Equation (63) is a very compact formula for calculating the cross spectra of the  $n \times 1$   $x_t$  process as a function of the fundamental parameters, the covariance matrix  $V$  and the coefficients in  $C(L)$  (or  $B(L)$ ). Equation (61) is quite a general representation and is flexible enough to incorporate *exogenous variables* and *serially correlated noises*.

In Equation (61) 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 (61) 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  $\varepsilon_{it}$ . In this sense the evolution of  $x_{it}$  is not affected by interactions with other variables in  $x_t$ . This is *not* to say however that  $x_{it}$  is uncorrelated with other components of  $x_t$  since  $\varepsilon_{it}$  can be correlated contemporaneously with other  $\varepsilon$ 's (i.e.,  $V$  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 18).

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 are not affected by other components of  $x_t$ .

### A. A Compact Notation

It is always possible to write an  $m$ th-order difference equation in terms of a vector first-order system. For example, consider the bivariate system

$$\begin{aligned} x_{1,t+1} &= \alpha_1 x_{1,t} + \cdots + \alpha_m x_{1,t-m+1} + \alpha_{m+1} x_{2,t} + \cdots \\ &\quad + \alpha_{2m} x_{2,t-m+1} + \varepsilon_{1,t+1} \\ x_{2,t+1} &= \beta_1 x_{1,t} + \cdots + \beta_m x_{1,t-m+1} + \beta_{m+1} x_{2,t} + \cdots \\ &\quad + \beta_{2m} x_{2,t-m+1} + \varepsilon_{2,t+1} \end{aligned} \quad (64)$$

where  $(\varepsilon_{1,t+1}, \varepsilon_{2,t+1})$  are two serially uncorrelated white-noise processes. Equations (64) can be written as

$$x_{t+1} = Ax_t + \varepsilon_{t+1} \quad (65)$$

where

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_m & \alpha_{m+1} & \cdots & \alpha_{2m} \\ 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & & & & \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ \beta_1 & \beta_2 & \cdots & \beta_m & \beta_{m+1} & \cdots & \beta_{2m} \\ 0 & 0 & & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & & \\ 0 & 0 & & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \leftarrow (m+1)\text{th row}$$

↑  $(m+1)$ th column

$$x_{t+1} = \begin{bmatrix} x_{1,t+1} \\ x_{1,t} \\ \vdots \\ x_{1,t-m+2} \\ x_{2,t+1} \\ x_{2,t} \\ \vdots \\ x_{2,t-m+2} \end{bmatrix}, \quad \varepsilon_{t+1} = \begin{bmatrix} \varepsilon_{1,t+1} \\ 0 \\ \vdots \\ 0 \\ \varepsilon_{2,t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow (m+1)\text{th row}$$

The solution of the vector difference equation (65) can be written

$$x_{t+\tau} = A^\tau x(t) + \varepsilon(t+\tau) + A\varepsilon(t+\tau-1) + \cdots + A^{\tau-1}\varepsilon(t+1). \quad (66)$$

Since  $E\varepsilon(t+\tau)x(t)' = 0_{2m \times 2m}$  for all  $\tau \geq 1$ , multiplying the solution (66) through by  $x(t)'$  and taking expected values gives the matrix Yule-Walker equation

$$Ex_{t+\tau}x_t' = A^\tau Ex_t x_t', \quad \tau \geq 1, \quad \text{or} \quad C_x(\tau) = A^\tau C_x(0), \quad \tau \geq 1, \quad (67)$$

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 (65), it is straightforward to show that the cross-spectral density matrix  $\Omega(e^{-i\omega})$  of the vector  $x$  process is given by

$$\Omega(e^{-i\omega}) = (e^{i\omega}I - A)^{-1}V(Ie^{-i\omega} - A')^{-1} \quad (68)$$

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  $\Lambda$  is the diagonal matrix whose diagonal entries are the eigenvalues of  $A$ . Then we have

$$A^r = P\Lambda^r P^{-1},$$

so that the solution (68) 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 complicated covariograms. Reference to this point can be used to show, for example, that while a one-variable, first-order difference equation cannot 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.

Formula (63) or (68) has been used to summarize and analyze the stochastic properties of linear macroeconomic models. For interesting examples of such work, the reader is referred to articles by Chow and Levitan (1969) and by Howrey (1971).

### B. Optimal Prediction: Compact Notation

Using the fact that  $\varepsilon_t$  in (65) is a serially uncorrelated vector process, it is straightforward to deduce from (66) that the projection of  $x_{t+\tau}$  against  $x_t$  is given by

$$P[x_{t+\tau}|x_t] = A^\tau x_t. \quad (69)$$

This is a compact formula for linear least squares predictors of a vector governed by a finite order stochastic difference equation.

Assuming that the eigenvalues of  $A$  are distinct so that  $A = P\Lambda P^{-1}$  where  $\Lambda$  is the diagonal matrix whose diagonal entries are the eigenvalues of  $A$ , we can write

$$P[x_{t+\tau}|x_t] = P\Lambda^\tau P^{-1}x_t. \quad (70)$$

As an example illustrating the use of this formula, return to the portfolio balance example (57), which leads to a solution for the price level of the form

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

where  $m_t$  is the log of the money supply. Suppose that  $m_t$  follows the second-order Markov process

$$m_t = \omega_0 + \omega_1 m_{t-1} + \omega_2 m_{t-2} + \zeta_t$$

where  $P_{t-1}\zeta_t = 0$ . Define

$$x_t = \begin{bmatrix} m_t \\ m_{t-1} \\ 1 \end{bmatrix}, \quad A = \begin{bmatrix} \omega_1 & \omega_2 & \omega_0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \varepsilon_t = \begin{bmatrix} \zeta_t \\ 0 \\ 0 \end{bmatrix},$$

so that  $x_t = Ax_{t-1} + \varepsilon_t$ . Let  $c$  be the row vector  $(1, 0, 0, \dots, 0)$ , so that  $m_t = cx_t$ . Then substituting (70) into the solution for  $p_t$  gives

$$\begin{aligned} p_t &= \frac{1}{1-\alpha} \sum_{j=0}^{\infty} \left( \frac{\alpha}{1-\alpha} \right)^j c P \Lambda^j P^{-1} x_t \\ &= c \frac{1}{1-\alpha} P \left[ \frac{1}{1 - (\alpha/(1-\alpha))\lambda_i} \right]_{ii} P^{-1} x_t \end{aligned}$$

where the matrix in brackets is diagonal and has  $(1 - (\alpha/(1-\alpha))\lambda_i)^{-1}$  as the  $i$ th element. The above derivation assumes that  $\max_i |(\alpha/(1-\alpha))\lambda_i| < 1$ , which is guaranteed<sup>28</sup> if  $\max_i |\lambda_i| < 1$ . The above formula is a compact representation of the solution for  $p_t$ . Notice that  $p_t$  depends on as many lags of  $m$  as the order of the Markov process for  $m$ .

### 17. 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  $y_t$  and  $x_t$  have means of zero and are jointly covariance stationary, indeterministic processes. That is, we seek the  $h_j$  that characterize the one-sided projection

$$y_t = \sum_{j=0}^{\infty} h_j x_{t-j} + u_t \quad (71)$$

where  $Ex_{t-j}u_t = 0$  for all  $j \geq 0$ . First, suppose that  $x_t$  has the moving average representation

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

<sup>28</sup> Nonstationarity of  $\{m_t\}$  in the form of  $\max |\lambda_i| > 1$  still leaves the proposed solution valid as long as  $\max_i |(\alpha/(1-\alpha))\lambda_i| < 1$ .

where  $\{\varepsilon_t\}$  is a serially uncorrelated process of innovations in  $x$ , i.e.,  $\varepsilon_t$  is fundamental for  $x$ . As an intermediate step,<sup>29</sup> think of projecting  $y_t$  on current and past  $\varepsilon$ 's:

$$y_t = \sum_{j=0}^{\infty} \phi_j \varepsilon_{t-j} + u_t \quad (72)$$

where  $E u_t \varepsilon_{t-j} = 0$  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  $\{x_t, x_{t-1}, \dots\}$  and  $\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$  span the same space. For this reason,  $u_t$  in (71) equals  $u_t$  in (72). Since the  $\varepsilon$ 's form an orthogonal process, we have that the  $\phi_j$  are the simple least squares coefficients

$$\phi_j = E y_t \varepsilon_{t-j} / E \varepsilon_t^2 = E y_t \varepsilon_{t-j} / \sigma^2$$

where  $\sigma^2 = E \varepsilon_t^2$ . Thus we can write

$$\phi(L) = \sum_{j=0}^{\infty} \phi_j L^j, \quad \phi(L) = \sigma^{-2} [g_{yx}(L)]_+ \quad (73)$$

where  $[\ ]_+$  again means "ignore negative powers of  $L$ " and  $g_{yx}(L)$  is the cross-covariance generating function

$$g_{yx}(L) = \sum_{k=-\infty}^{\infty} E(y_t \varepsilon_{t-k}) L^k.$$

We can relate  $g_{yx}(L)$  to the cross-covariance generating function  $g_{yx}(L)$  as follows:

$$\begin{aligned} g_{yx}(z) &= \sum_k (E y_t x_{t-k}) z^k \\ &= \sum_k (E y_t d(L) \varepsilon_{t-k}) z^k \\ &= \sum_k (E y_t (d_0 \varepsilon_{t-k} + d_1 \varepsilon_{t-k-1} + \dots)) z^k \\ &= d_0 \sum_k (E y_t \varepsilon_{t-k}) z^k + d_1 \sum_k (E y_t \varepsilon_{t-k-1}) z^k + d_2 \sum_k (E y_t \varepsilon_{t-k-2}) z^k + \dots \\ &= d_0 g_{yx}(z) + d_1 z^{-1} g_{yx}(z) + d_2 z^{-2} g_{yx}(z) + \dots \\ &= d(z^{-1}) g_{yx}(z). \end{aligned}$$

Thus we have  $g_{yx}(z) = g_{yx}(z)/d(z^{-1})$ . Substituting this into (73), we obtain

$$\phi(L) = \frac{1}{\sigma^2} \left( \frac{g_{yx}(L)}{d(L^{-1})} \right)_+.$$

<sup>29</sup> This is the method that Kolmogorov used to derive the formula we are after. See Whittle (1963, p. 42).

So we have

$$\begin{aligned} y_t &= \frac{1}{\sigma^2} \left( \frac{g_{yx}(L)}{d(L^{-1})} \right)_+ \varepsilon_t + u_t \\ &= \frac{1}{\sigma^2} \left( \frac{g_{yx}(L)}{d(L^{-1})} \right)_+ \frac{1}{d(L)} x_t + u_t, \end{aligned} \quad (74)$$

so that in (71) we have

$$h(L) = \frac{1}{\sigma^2} \left( \frac{g_{yx}(L)}{d(L^{-1})} \right)_+ \frac{1}{d(L)}. \quad (75)$$

A classic application of this formula is due to Muth (1960). Suppose that income evolves according to  $x_t = y_t + \varepsilon_t$  where  $y_t = \rho y_{t-1} + u_t$ ,  $|\rho| < 1$ , and where  $u_t$  and  $\varepsilon_t$  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"  $x_t, x_{t-1}, \dots$  and desires to estimate systematic income  $y_t$  by a linear function of  $x_t, x_{t-1}, \dots$ . The consumer is assumed to know all the relevant moments. This problem can be solved quickly using formula (75), and the reader is invited to do so. A more tedious method of solution is adopted in Chapter XII.

## 18. THE RELATIONSHIP BETWEEN WIENER-GRANGER CAUSALITY AND ECONOMETRIC EXOGENEITY

We now study the conditions under which the projection of  $y_t$  on the entire  $x$  process equals the projection of  $y_t$  on only current and all past  $x$ 's, i.e., the condition under which

$$g_{yx}(z)/g_x(z) = [g_{yx}(z)/d(z^{-1})]_+ (1/d(z)\sigma^2)$$

where  $g_x(z) = \sigma^2 d(z) d(z^{-1})$ . Sims (1972a) proved the important result that these two projections are equal if and only if lagged  $y$ 's fail linearly to help predict  $x$ , given lagged  $x$ 's. We shall first prove Sims's result in a different way than he did. The method is in the spirit of Wiener's derivation of the Wiener-Kolmogorov prediction formula and has certain advantages when it comes to working Exercises 6-13.

We consider a jointly covariance stationary stochastic process  $\{y_t, x_t\}$ , with  $E x_t = E y_t = 0$  and with covariance generating functions  $g_x(z), g_y(z), g_{yx}(z)$ . We assume that  $x$  possesses an autoregressive representation and that both  $y$  and  $x$  are linearly indeterministic. We now consider the projection of  $x_t$  on past values of  $x$  and past values of  $y$ :

$$x_t = \sum_{j=1}^{\infty} h_j x_{t-j} + \sum_{j=1}^{\infty} v_j y_{t-j} + u_t \quad (76)$$



where the least squares residual  $u_t$  obeys the orthogonality conditions  $Eu_t x_{t-\tau} = Eu_t y_{t-\tau} = 0$  for  $\tau = 1, 2, \dots$ . Solving (76) for  $u_t$  permits the orthogonality conditions to assume the form of the normal equations

$$E\left\{x_t - \sum_{j=1}^{\infty} h_j x_{t-j} - \sum_{j=1}^{\infty} v_j y_{t-j}\right\} x_{t-\tau} = 0, \quad \tau = 1, 2, \dots$$

$$E\left\{x_t - \sum_{j=1}^{\infty} h_j x_{t-j} - \sum_{j=1}^{\infty} v_j y_{t-j}\right\} y_{t-\tau} = 0, \quad \tau = 1, 2, \dots$$

These equations can be written

$$c_x(\tau) = \sum_{j=1}^{\infty} h_j c_x(\tau - j) + \sum_{j=1}^{\infty} v_j c_{yx}(\tau - j), \quad (77)$$

$$c_{xy}(\tau) = \sum_{j=1}^{\infty} h_j c_{xy}(\tau - j) + \sum_{j=1}^{\infty} v_j c_y(\tau - j), \quad (78)$$

which are required to hold *only for positive integers*  $\tau = 1, 2, \dots$ . Multiplying both sides of (77) and (78) by  $z^\tau$  and summing over *all*  $\tau$ , we get the following equation in terms of the  $z$  transforms

$$g_x(z) + m(z) = h(z)g_x(z) + v(z)g_{yx}(z), \quad (79)$$

$$g_{xy}(z) + n(z) = h(z)g_{xy}(z) + v(z)g_y(z), \quad (80)$$

where  $m(z)$  and  $n(z)$  are each unknown series in *nonpositive powers* of  $z$  only. That  $m(z)$  and  $n(z)$  are series in *nonpositive powers* of  $z$  is equivalent with Equations (77) and (78) holding only for  $\tau \geq 1$ . Equations (79) and (80) are the normal equations for  $h(z)$  and  $v(z)$ .

Following Wiener, Granger (1969) has proposed the terminology that " $y$  causes  $x$ " whenever  $v(z) \neq 0$ . That is,  $y$  is said to cause  $x$  if, given all past values of  $x$ , past values of  $y$  help to predict  $x$ . The conditions under which  $v(z)$  does or does not equal zero turn out to be of substantial interest to econometricians and macroeconomists, which is the reason that this concept of causality is an interesting one to study.

Consider the projection of  $y_t$  on the entire  $x$  process,

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

where  $E\varepsilon_t x_{t-j} = 0$  for all  $j$ . Under what conditions will the lag distribution  $\{b_j\}$  be one-sided on the past and present, so that  $b_j = 0$ , for  $j < 0$ ? From formula (38) we have that

$$b(z) = g_{yx}(z)/g_x(z). \quad (81)$$

Suppose that  $x_t$  has the Wold moving average representation

$$x_t = d(L)\eta_t, \quad \eta_t = x_t - P[x_t | x_{t-1}, \dots], \quad \sum_{j=0}^{\infty} d_j^2 < \infty.$$

Then

$$g_x(z) = \sigma_\eta^2 d(z) d(z^{-1}). \quad (82)$$

We have assumed that  $x$  possesses an autoregressive representation so that  $[d(z)]^{-1}$  is one-sided and square summable in nonnegative powers of  $z$ . Now it is always possible uniquely to factor the cross-covariance generating function as

$$g_{yx}(z) = a(z)\phi(z^{-1}) \quad (83)$$

where both  $a(z)$  and  $\phi(z)$  are one-sided in nonnegative powers of  $z$ .<sup>30</sup> Substituting (83) and (82) into (81) gives

$$b(z) = a(z)\phi(z^{-1})/\sigma_\eta^2 d(z) d(z^{-1}).$$

Evidently,  $b(z)$  is one-sided in nonnegative powers of  $z$  if and only if  $\phi(z^{-1}) = k d(z^{-1})$ , where  $k$  is a constant.<sup>31</sup> Under this condition (81) becomes  $b(z) = ka(z)/\sigma_\eta^2 d(z)$ . We shall assume that  $a(z)$  has an inverse that is one-sided in nonnegative powers of  $z$ .

We can now prove the following important theorem due to Sims.

**Theorem:**  $v(z) = 0$  if and only if  $b(z)$  is one-sided in nonnegative powers of  $z$ .

*Proof:* Suppose that  $b(z)$  is one-sided in nonnegative powers of  $z$ . Then we know that  $g_{yx}(z) = ka(z) d(z^{-1})$ . We must show that (79) and (80) are satisfied with  $v(z) = 0$ . Now if  $v(z) = 0$ , then (79) becomes

$$\sigma_\eta^2 d(z) d(z^{-1}) + m(z) = h(z)\sigma_\eta^2 d(z) d(z^{-1}).$$

Dividing both sides by  $z$  gives

$$\sigma_\eta^2 \frac{d(z) d(z^{-1})}{z} + \frac{m(z)}{z} = \frac{h(z)}{z} \sigma_\eta^2 d(z) d(z^{-1})$$

where  $m(z)/z$  is now a series in strictly negative powers of  $z$ . Dividing both sides by  $\sigma_\eta^2 d(z^{-1})$  gives

$$\frac{d(z)}{z} + \frac{m(z)}{\sigma_\eta^2 z d(z^{-1})} = \frac{h(z)}{z} d(z).$$

<sup>30</sup> This can be proved by using the method of Whittle (1963, p. 26).

<sup>31</sup> From the optimum filtering formula (74) we have that the  $z$  transform  $f(z)$  of the coefficients  $f_j$  in the projection of  $y$  on current and past  $x$ 's is given by

$$f(z) = \frac{1}{\sigma_\eta^2} \left[ \frac{a(z)\phi(z^{-1})}{d(z^{-1})} \right]_+ \frac{1}{d(z)}.$$

Evidently,  $f(z) = b(z)$  if and only if  $\phi(z^{-1}) = kd(z^{-1})$ .

The term  $m(z)/z d(z^{-1})$  involves only negative powers of  $z$ , while the term  $h(z) d(z)/z$  involves only nonnegative powers of  $z$ . Therefore, we have

$$\left[ \frac{d(z)}{z} \right]_+ = \frac{h(z)}{z} d(z) \quad \text{or} \quad z \left[ \frac{d(z)}{z} \right]_+ \frac{1}{d(z)} = h(z), \quad (84)$$

which is just the Wiener-Kolmogorov formula. In fact, the foregoing is Wiener's derivation of that formula. Now if  $v(z) = 0$  and  $g_{yx}(z) = ka(z) d(z^{-1})$ , Equation (80) becomes

$$ka(z^{-1}) d(z) + n(z) = h(z)ka(z^{-1}) d(z).$$

This can be rewritten as

$$\frac{d(z)}{z} + \frac{n(z)}{zka(z^{-1})} = \frac{h(z)}{z} d(z). \quad (85)$$

Again, since  $n(z)/zka(z^{-1})$  involves only strictly negative powers of  $z$ , the solution of this equation is the Wiener-Kolmogorov formula (52). Therefore, if  $b(z)$  is one-sided in nonnegative powers of  $z$ , the normal equations (79) and (80) are both satisfied with  $v(z) = 0$  and

$$h(z) = z[d(z)/z]_+ d(z)^{-1}.$$

Now suppose that the normal equations (79) and (80) are satisfied with  $v(z) = 0$ . Then Equation (80) becomes

$$\phi(z)a(z^{-1}) + n(z) = z \left[ \frac{d(z)}{z} \right]_+ \frac{1}{d(z)} a(z^{-1})\phi(z).$$

Dividing both sides by  $za(z^{-1})$  gives

$$\frac{\phi(z)}{z} + \frac{n(z)}{za(z^{-1})} = \left[ \frac{d(z)}{z} \right]_+ \frac{\phi(z)}{d(z)} \quad (86)$$

where  $n(z)/za(z^{-1})$  involves only negative powers of  $z$ . Since the right-hand side involves only nonnegative powers of  $z$ , applying  $[ ]_+$  leaves the right-hand side unaltered, so that (86) implies

$$d(z)[\phi(z)/z]_+ = [d(z)/z]_+ \phi(z).$$

This equation can be satisfied only if  $\phi(z) = k d(z)$ , where  $k$  is a constant. This completes the proof.

In words, Sims's theorem asserts that the projection of  $y$  on the entire  $x$  process equals the projection of  $y$  on current and past  $x$ 's if and only if  $y$  fails to Granger cause  $x$  (i.e.,  $y$  fails to help predict  $x$ ).

**Example:** Suppose that

$$g_x(z) = (1 + 0.4z)(1 + 0.4z^{-1}), \quad g_y(z) = (1 - 0.2z)(1 - 0.2z^{-1}),$$

$$g_{yx}(z) = \frac{(1 - 0.2z)}{1 - 0.9z^{-1}}.$$

The projection of  $y$  on the entire  $x$  process has coefficient generating function

$$\frac{g_{yx}(z)}{g_x(z)} = \frac{1 - 0.2z}{(1 - 0.9z^{-1})(1 + 0.4z)(1 + 0.4z^{-1})},$$

which has nonzero coefficients on negative powers of  $z$  (expand the polynomial in  $z$  by partial fractions to verify this). Therefore  $y$  Granger causes (helps predict)  $x$ . The projection of  $x$  on the entire  $y$  process has the coefficient generating function

$$\frac{g_{xy}(z)}{g_y(z)} = \frac{g_{yx}(z^{-1})}{g_y(z)} = \frac{1 - 0.2z^{-1}}{(1 - 0.9z)(1 - 0.2z^{-1})(1 - 0.2z)}$$

$$= \frac{1}{(1 - 0.9z)(1 - 0.2z)},$$

which involves only nonnegative powers of  $z$ . Therefore  $x$  fails to Granger cause  $y$  (i.e., given lagged  $y$ 's,  $x$  fails to help predict  $y$ ).

To gain additional perspective on Sims's theorem, we now undertake to indicate the proof that Sims gave. We do this not only because the theorem is very important, but because his proof provides useful practice in projection arguments and useful insights into the nature of bivariate Wold representations.

Let  $[x_t^*]$  be a bivariate, jointly covariance stationary stochastic process. Suppose that  $[x_t^*]$  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

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} c^{11}(L) & c^{12}(L) \\ c^{21}(L) & c^{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

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  $L$ ;  $\varepsilon_t$  and  $u_t$  are serially uncorrelated processes with  $E u_t \varepsilon_s = 0$  for all  $t, s$ ;  $E \varepsilon_t^2 = \sigma_\varepsilon^2$ ,  $E u_t^2 = \sigma_u^2$ ; and where the one-step-ahead prediction errors are given by

$$x_t - P[x_t | x_{t-1}, \dots, y_{t-1}, \dots] = c_0^{11} \varepsilon_t + c_0^{12} u_t,$$

$$y_t - P[y_t | x_{t-1}, \dots, y_{t-1}, \dots] = c_0^{21} \varepsilon_t + c_0^{22} u_t,$$

i.e.,  $\varepsilon$  and  $u$  are "jointly fundamental for  $x$  and  $y$ ."<sup>32</sup> 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 shall 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

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = F_1^n \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \dots + F_n^n \begin{bmatrix} x_{t-n} \\ y_{t-n} \end{bmatrix} + \begin{bmatrix} a_{xt}^n \\ a_{yt}^n \end{bmatrix} \quad (87)$$

where  $F_1^n, \dots, F_n^n$  are  $2 \times 2$  matrices of least squares coefficients and we have the orthogonality conditions

$$E \begin{bmatrix} x_{t-j} \\ y_{t-j} \end{bmatrix} \begin{bmatrix} a_{xt}^n & a_{yt}^n \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

for  $j = 1, \dots, n$ . We assume that as  $n \rightarrow \infty$ , the  $F_j^n$  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

$$\begin{aligned} \begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \sum_{j=1}^{\infty} F_j \begin{bmatrix} x_{t-j} \\ y_{t-j} \end{bmatrix} + \begin{bmatrix} a_{xt} \\ a_{yt} \end{bmatrix} \\ &= F(L) \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} a_{xt} \\ a_{yt} \end{bmatrix}, \quad F(L) = \sum_{j=1}^{\infty} F_j L^{j-1} \end{aligned}$$

where the random variables  $(a_{xt}, a_{yt})$  obey the least squares orthogonality conditions

$$E \begin{bmatrix} x_{t-j} \\ y_{t-j} \end{bmatrix} \begin{bmatrix} a_{xt} & a_{yt} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

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:

$$A(L) \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \quad \text{or} \quad (A_0 - A_1 L - A_2 L^2 \dots) \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \quad (88)$$

<sup>32</sup> The statement " $\{\varepsilon_t\}$  and  $\{u_t\}$  are jointly fundamental for  $\{x_t, y_t\}$ " means that the one-step-ahead errors in forecasting  $(y_t, x_t)$  from past  $x$ 's and  $y$ 's are linear combinations of  $\varepsilon_t$  and  $u_t$ .

where  $A_j$  is a  $2 \times 2$  matrix for each  $j$ , where  $A_0$  is chosen to be lower triangular and  $\{\varepsilon_t\}$  and  $\{u_t\}$  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  $\varepsilon$  and  $u$  being orthogonal processes? The answer is in general yes,<sup>33</sup> as the following argument suggests. Think of projecting  $x_t$  against all *lagged*  $x$ 's and *lagged*  $y$ 's. This gives the first row of  $A(L)$  and gives a least squares residual process  $\varepsilon_t$  that is by construction orthogonal to all lagged  $y$ 's and all lagged  $x$ 's. Next project  $y_t$  against *current* 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 *current* and *lagged*  $x$ 's and lagged  $y$ 's. This procedure produces an  $A_0$  that is lower triangular as required. Further, notice that since  $\varepsilon_t$  is orthogonal to all lagged  $x$ 's and  $y$ 's and since the representation (88) that we have achieved permits lagged  $\varepsilon$ 's and  $u$ 's to be expressed as linear combinations of lagged  $x$ 's and  $y$ 's, it follows that  $\varepsilon_t$  is orthogonal to lagged  $u$ 's and  $\varepsilon$ 's. A similar argument shows that  $u_t$  is orthogonal to lagged  $u$ 's and  $\varepsilon$ 's. Finally, since by construction  $u_t$  is orthogonal to current and lagged  $x$ 's and lagged  $y$ 's and since  $\varepsilon_t$  is by definition a linear combination of current and lagged  $x$ 's and lagged  $y$ 's, it follows that  $u_t$  and  $\varepsilon_t$  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  $\begin{bmatrix} u_t \\ \varepsilon_t \end{bmatrix}$  that satisfies the same conditions on second moments that the  $\begin{bmatrix} x_t \\ y_t \end{bmatrix}$  process satisfies.

To get (88) in a form that is useful for studying prediction problems, pre-multiply (88) by  $A_0^{-1}$  to get<sup>34</sup>

$$A_0^{-1} A(L) \begin{bmatrix} x_t \\ y_t \end{bmatrix} = A_0^{-1} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

<sup>33</sup> 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  $A(e^{-i\omega}) = C(e^{-i\omega})^{-1}$  for each  $\omega$  between  $-\pi$  and  $\pi$ . The autoregressive representation exists provided that  $C(e^{-i\omega})$  is invertible at each frequency between  $-\pi$  and  $\pi$ . This condition is a restriction but is one that can usually be assumed in applied work. (For an example of a  $C(e^{-i\omega})$  that violates the condition, consider the univariate example  $C(e^{-i\omega}) = 1 - e^{-i\omega}$ —the transform of the first difference operator  $1 - L$ —which equals zero at  $\omega = 0$  and so is not invertible there.)

<sup>34</sup> Notice that (89) is identical with (87) for  $n = \infty$ , so that we must have  $F_j = A_0^{-1} A_j$ ,

$$\begin{bmatrix} a_{xt} \\ a_{yt} \end{bmatrix} = A_0^{-1} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}.$$

Notice that  $(a_{xt}, a_{yt})$  are by the orthogonality conditions serially uncorrelated and uncorrelated with one another at all nonzero lags.

or

$$\begin{aligned} \begin{bmatrix} x_t \\ y_t \end{bmatrix} &= A_0^{-1} [A_1 L + A_2 L^2 + \dots] \begin{bmatrix} x_t \\ y_t \end{bmatrix} + A_0^{-1} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \\ &= A_0^{-1} H(L) \begin{bmatrix} x_t \\ y_t \end{bmatrix} + A_0^{-1} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \end{aligned} \quad (89)$$

where  $H(L) = A_1 L + A_2 L^2 + \dots$ . The linear least squares prediction of the  $\begin{bmatrix} x_t \\ y_t \end{bmatrix}$  process based on all lagged  $x$ 's and all lagged  $y$ 's (call it  $P_{t-1} \begin{bmatrix} x_t \\ y_t \end{bmatrix}$ ) from (89) is then

$$P_{t-1} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = A_0^{-1} H(L) \begin{bmatrix} x_t \\ y_t \end{bmatrix} = F(L) L \begin{bmatrix} x_t \\ y_t \end{bmatrix}$$

since by construction  $P_{t-1} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} = 0$ . The one-step-ahead prediction errors in predicting the  $\begin{bmatrix} x_t \\ y_t \end{bmatrix}$  process are given by

$$A_0^{-1} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}.$$

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  $\varepsilon_t$  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  $\varepsilon_t$  process was constructed above.

If  $A_0^{-1} A(L)$  is lower triangular (i.e., the matrix coefficient is lower triangular for each power of  $L$ ), then given lagged  $x$ 's, lagged  $y$ 's do not 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 Wiener and Granger,  $y$  is said to *cause*  $x$  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 *failing* to cause  $x$ , in the Wiener-Granger sense.

Given that  $A_0^{-1}$  is lower triangular, 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

$$A(L)^{-1} = A(L)^{-1} A_0 A_0^{-1}.$$

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.<sup>35</sup>

<sup>35</sup> To make the argument in terms of ordinary matrices, write  $A(e^{-i\omega})^{-1} = A(e^{-i\omega})^{-1} A_0 A_0^{-1}$  and note that  $A(e^{-i\omega})^{-1} A_0$  is the inverse of the lower triangular matrix  $A_0^{-1} A(e^{-i\omega})$  at each frequency and so is lower triangular. It follows that  $A(e^{-i\omega})^{-1}$  is lower triangular (at each frequency) being the product of two lower triangular matrices. It then follows that

$$A_j^{-1} = (1/2\pi) \int_{-\pi}^{\pi} A(e^{-i\omega})^{-1} e^{i\omega j} d\omega$$

is lower triangular for  $j = 0, 1, 2, \dots$

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 (88) can be "inverted" to yield the vector moving average representation

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = C(L) \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \quad (90)$$

where  $A(L)^{-1} = C(L) = C_0 + C_1 L + C_2 L^2 + \dots$ ,  $C_j$  being a  $2 \times 2$  matrix, and where  $C(L)$  is lower triangular. Recall the extensive orthogonality conditions satisfied by  $\varepsilon$  and  $u$ : the  $\varepsilon$  and  $u$  processes are orthogonal at all lags, even contemporaneously.<sup>36</sup> Conversely, suppose that a moving average representation of the lower triangular form (90) exists with  $\varepsilon_t$  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

$$C(L)^{-1} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \quad \text{or} \quad A(L) \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

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 proved 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

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} C^{11}(L) & 0 \\ C^{21}(L) & C^{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

<sup>36</sup> Assuming that things have been normalized so that  $\varepsilon$  and  $u$  have unit variances, the spectral density matrix of the  $(x, y)$  process satisfying (90) is, as we have seen,

$$S(e^{-i\omega}) = C(e^{-i\omega}) I C(e^{-i\omega})'$$

where the prime now denotes both complex conjugation and transposition. Now let  $U$  be a  $2 \times 2$  unitary matrix, i.e., a  $2 \times 2$  matrix satisfying  $UU' = U'U = I$  where here the prime again denotes complex conjugation and transposition. Then note that  $S(e^{-i\omega})$  can also be represented

$$\begin{aligned} S(e^{-i\omega}) &= C(e^{-i\omega}) U U' C(e^{-i\omega})' = [C(e^{-i\omega}) U] I [C(e^{-i\omega}) U]' \\ &= D(e^{-i\omega}) I D(e^{-i\omega})' \end{aligned}$$

where  $D(e^{-i\omega}) = C(e^{-i\omega}) U$ . 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  $C(e^{-i\omega})$  by  $U$  will, in general, destroy the lower triangularity of  $C(e^{-i\omega})$  if  $C$  originally has this property.

where  $\varepsilon_t$  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  $(x_t - P[x_t | x_{t-1}, \dots, y_{t-1}, \dots])$  and  $(y_t - P[y_t | x_{t-1}, \dots, y_{t-1}, \dots])$  are each linear combinations of  $\varepsilon_t$  and  $u_t$  (Sims, 1972a).

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  $x$  (Sims, 1972a).

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 it 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 (90) is lower triangular, so that

$$x_t = C^{11}(L)\varepsilon_t, \quad (91)$$

$$y_t = C^{21}(L)\varepsilon_t + C^{22}(L)u_t, \quad (92)$$

where all polynomials in  $L$  involve only nonnegative powers of  $L$ . Inverting (91) and substituting into (92) gives<sup>37</sup>

$$y_t = C^{21}(L)C^{11}(L)^{-1}x_t + C^{22}(L)u_t$$

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  $\varepsilon_t$  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  $x_t$

$$y_t = h(L)x_t + \eta_t, \quad x_t = a(L)\varepsilon_t,$$

<sup>37</sup> In assuming that  $(x_t, y_t)$  has an autoregressive representation we have in effect assumed that  $C^{11}(L)$  has an inverse that is one-sided in nonnegative powers of  $L$ .

where by hypothesis  $\eta$  is orthogonal to  $x$  and therefore to  $\varepsilon$  at all lags. Then by finding the moving average representation for  $\eta_t$ , 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

$$y_t = h(L)a(L)\varepsilon_t + m(L)u_t, \quad x_t = a(L)\varepsilon_t$$

or

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = C(L) \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

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 (90), 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

$$[I - A_0^{-1}A_1L - A_0^{-1}A_2L^2 - \dots] \begin{bmatrix} x_t \\ y_t \end{bmatrix} = A_0^{-1} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}.$$

The lower triangularity of the matrices on the left and the orthogonality properties of  $\varepsilon$  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.

## 19. SIMS'S APPLICATION TO MONEY AND INCOME

Economists at the Federal Reserve Bank of St. Louis (Andersen and Jordan, 1968) have computed estimates of one-sided distributed lag regressions of (the log of) nominal income ( $y_t$ ) against (the log of) money ( $m_t$ ):

$$y_t = \sum_{j=0}^{\infty} h_j m_{t-j} + \eta_t, \quad (93)$$

where  $E\eta_t m_{t-j} = 0$  for  $j = 0, 1, 2, \dots$ . Those economists recommend that the  $h_j$  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  $h_j$  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 macroeconomic 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 (93), e.g.,

$$y_t = \sum_{j=0}^{\infty} v_j m_{t-j} + \sum_{j=0}^{\infty} w_j z_{t-j} + \varepsilon_t \quad (94)$$

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  $\varepsilon_t$  is a stationary stochastic process that obeys the orthogonality conditions  $E\varepsilon_t m_{t-j} = E\varepsilon_t z_{t-j} = 0$  for  $j = 0, \pm 1, \pm 2, \dots$ .

The strong condition that  $\varepsilon$  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 (94). These orthogonality conditions characterize (94) as a "final form" relationship. In (94) the  $v_j$  are the dynamic money multipliers and depict the average response of  $y_t$  to a unit impulse in  $m_t$  holding constant the  $z$ 's. Applying the law of iterated projections to (94), we obtain

$$P[y_t | m_t, m_{t-1}, \dots] = \sum_{j=0}^{\infty} v_j m_{t-j} + \sum_{k=0}^{\infty} w_k P[z_{t-k} | m_t, m_{t-1}, \dots].$$

Let  $P[z_{t-k} | m_t, m_{t-1}, \dots] = \sum_{j=0}^{\infty} \alpha_{kj} m_{t-j}$ . Then we have<sup>38</sup>

$$P[y_t | m_t, m_{t-1}, \dots] = \sum_{j=0}^{\infty} v_j m_{t-j} + \sum_{k=0}^{\infty} w_k \sum_{j=0}^{\infty} \alpha_{kj} m_{t-j}$$

or

$$y_t = \sum_{j=0}^{\infty} \left( v_j + \sum_{k=0}^{\infty} w_k \alpha_{kj} \right) m_{t-j} + \eta_t \quad (95)$$

where by the orthogonality principle we have  $E\eta_t m_{t-j} = 0, j = 0, 1, 2, \dots$ . Now (95) is identical with (93), so that the population  $h_j$  of (93) obey

$$h_j = v_j + \sum_{k=0}^{\infty} w_k \alpha_{kj}.$$

Therefore, the  $h_j$  in general *do not* equal the money multipliers, the  $v_j$ . The  $h_j$  are "mongrel" coefficients that do not indicate the typical average response of  $y$  to exogenous impulses in  $m_t$ , everything else, namely the  $z$ 's, being held constant. For this reason, Keynesians would argue, estimating Equation (93) is not a good way of estimating the dynamic multipliers, the  $v_j$ .

<sup>38</sup> This is a version of H. Theil's "omitted variable theorem." See Theil (1971, pp. 548-550).

Now project both sides of (94) against the entire sequence  $\{m_{t-j}\}_{j=-\infty}^{\infty}$  to get

$$y_t = \sum_{j=0}^{\infty} h_j m_{t-j} + \sum_{k=0}^{\infty} w_k \sum_{j=-\infty}^{\infty} \gamma_{kj} m_{t-j} + \xi_t \quad (96)$$

where  $E\xi_t m_{t-j} = 0$  for all  $j$  and

$$P(z_{t-k} | \{m_{t-j}\}_{j=-\infty}^{\infty}) = \sum_{j=-\infty}^{\infty} \gamma_{kj} m_{t-j}$$

where  $\gamma_{kj}$  is a vector of coefficients. We can write (96) as

$$P(y_t | \{m_{t-j}\}_{j=-\infty}^{\infty}) = \sum_{j=-\infty}^{\infty} d_j m_{t-j}$$

where

$$d_j = \begin{cases} h_j + \sum_{k=0}^{\infty} w_k \gamma_{kj}, & j \geq 0, \\ \sum_{k=0}^{\infty} w_k \gamma_{kj}, & j < 0. \end{cases}$$

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  $h_j$  are mongrel parameters), the  $\gamma_{kj}$  and therefore the  $d_j$  will not vanish for some  $j < 0$ . That is because in general future  $m$ 's will help explain current and past  $z_t$ .<sup>39</sup> Therefore, so long as the  $w_k$  are not zero in the final form (94), 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. For this reason, a test of the null hypothesis that the projection of  $y_t$  on the entire  $\{m\}$  process is one-sided (i.e., it equals the projection of  $y_t$  on current and past  $m$ 's alone) can be regarded as testing the null hypothesis that the  $w_k$  in (94) are zeros. But remember that the contention that the  $w_k$  are not zero is what underlies the Keynesian objection against interpreting the St. Louis equation's  $h_j$  as estimates of the dynamic money multipliers. So computing the two-sided projection

$$y_t = \sum_{j=-\infty}^{\infty} \delta_j m_{t-j} + \hat{\eta}_t \quad (97)$$

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"—i.e., that it is appropriate to set the  $w_k$  equal to zero.

Using post-World War II U.S. data, Sims estimated (97) 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 (97). In general, if the Keynesian objection to the St. Louis equation were correct, in large enough samples

<sup>39</sup> Unless  $m_t$  is strictly exogenous with respect to the vector  $z_t$ , or, equivalently, the vector  $z_t$  does not Granger cause  $m_t$ .

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 the 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 (93) 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 (93) 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 (93) 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  $\eta_t$  in (93) would not obey the restrictions  $E\eta_t m_{t-s} = 0$  for all  $s$ ; i.e., (93) 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

$$m_t = C^{11}(L)\varepsilon_t, \quad y_t = d(L)m_t + C^{22}(L)u_t$$

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 (but do not necessarily imply) 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.

## 20. MULTIVARIATE PREDICTION FORMULAS

Continue to assume that  $(x_t, y_t)$  is a jointly covariance stationary, strictly indeterministic process with a moving average representation

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} C^{11}(L) & C^{12}(L) \\ C^{21}(L) & C^{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} = C(L) \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

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

$$C(L)^{-1} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \quad \text{or} \quad A(L) \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

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  $\{x_t, x_{t-1}, \dots, y_t, y_{t-1}, \dots\}$ , call it  $P_t[\begin{smallmatrix} x_{t+1} \\ y_{t+1} \end{smallmatrix}]$ , is

$$P_t \begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} = \left( \frac{C(L)}{L} \right)_+ \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} = \left( \frac{C(L)}{L} \right)_+ A(L) \begin{bmatrix} x_t \\ y_t \end{bmatrix}.$$

More generally, we have

$$P_t \begin{bmatrix} x_{t+j} \\ y_{t+j} \end{bmatrix} = \left( \frac{C(L)}{L^j} \right)_+ A(L) \begin{bmatrix} x_t \\ y_t \end{bmatrix}$$

These results 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

$$y_t = C(L)\varepsilon_t \quad (98)$$

where  $C(L) = C_0 + C_1L + \dots$ ,  $C_j$  being an  $n \times n$  matrix and the  $C_j$  being "square summable," where  $\varepsilon_t$  is an  $(n \times 1)$ -vector stochastic process, where the component  $\varepsilon_{it}$  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  $\varepsilon_{it}$  are "jointly fundamental for  $y_t$ ," i.e., for each  $i$ ,  $y_{it} - (Py_{it}|y_{t-1}, y_{t-2}, \dots)$  is a linear combination of  $\varepsilon_{it}, j = 1, \dots, n$ . For the process (98), we have the prediction formula

$$E_t y_{t+j} = [C(L)/L^j]_+ \varepsilon_t$$

where  $E_t(x) \equiv Ex|y_t, y_{t-1}, \dots$ . Where  $C(L)^{-1}$  exists, so that  $y_t$  has a vector autoregressive representation, then we also have the formula

$$E_t y_{t+j} = [C(L)/L^j]_+ C(L)^{-1} y_t.$$

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

$$R_{nt} = \alpha(L)\varepsilon_t + \beta(L)u_t, \quad n > 1, \quad R_{1t} = \gamma(L)\varepsilon_t + \delta(L)u_t \quad (99)$$

where all lag operators are one-sided on the present and past, and

$$R_{nt} - P_{t-1}R_{nt} = \alpha_0\varepsilon_t + \beta_0u_t, \quad R_{1t} - P_{t-1}R_{1t} = \gamma_0\varepsilon_t + \delta_0u_t.$$

The rational expectations theory of the term structure asserts<sup>40</sup>

$$\begin{aligned} R_{nt} &= \frac{1}{n} [R_{1t} + P_t R_{1t+1} + \cdots + P_t R_{1t+n-1}] \\ &= \frac{1}{n} \left[ \gamma(L) + \frac{\gamma(L)}{L} + \cdots + \frac{\gamma(L)}{L^{n-1}} \right]_+ \varepsilon_t \\ &\quad + \frac{1}{n} \left[ \delta(L) + \frac{\delta(L)}{L} + \cdots + \frac{\delta(L)}{L^{n-1}} \right]_+ u_t \end{aligned}$$

or

$$R_{nt} = \frac{1}{n} \left[ \left[ \frac{1-L^{-n}}{1-L^{-1}} \right] \gamma(L) \right]_+ \varepsilon_t + \frac{1}{n} \left[ \left[ \frac{1-L^{-n}}{1-L^{-1}} \right] \delta(L) \right]_+ u_t. \quad (100)$$

Thus, comparing (99) with (100), 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:

$$\alpha(L) = \frac{1}{n} \left[ \left[ \frac{1-L^{-n}}{1-L^{-1}} \right] \gamma(L) \right]_+, \quad \beta(L) = \frac{1}{n} \left[ \left[ \frac{1-L^{-n}}{1-L^{-1}} \right] \delta(L) \right]_+.$$

These restrictions embody the content of the theory and are refutable.

### EXERCISES

1. (Sims's approximation error formula) Let  $(y_t, x_t)$  be jointly covariance stationary with means of zero. Let the projection of  $y_t$  on the  $x$  process be

$$\sum_{j=-\infty}^{\infty} b_j^0 x_{t-j}.$$

Suppose a researcher fits by least squares

$$y_t = \sum_{j=-\infty}^{\infty} b_j^1 x_{t-j} + u_t$$

where  $u_t$  is a disturbance and  $\{b_j^1\}$  is a constrained parametrization so that  $b_j^1$  cannot equal  $b_j^0$  for all  $j$ . Some examples of commonly encountered constrained parametrizations are:

- (i) truncation:  $b_j^1 = 0$  for  $|j| > m$ ,  $m$  a fixed positive integer;
- (ii) polynomial approximation:  $b_j^1 = \alpha_0 + \alpha_1 j + \cdots + \alpha_m j^m$ ,  $m$  a fixed positive integer,  $\alpha_j$  free;
- (iii) Pascal lag distributions (Solow)

$$b^1(L) = \frac{1}{(1-\lambda L)^r}$$

where  $r$  is a fixed positive integer and  $|\lambda| < 1$ .

<sup>40</sup> Assuming that information used to forecast  $R_{1t}$  is confined to current and past  $R_{1t}$  and  $R_{nt}$  alone.

Derive Sims's formula, which asserts that in population, least squares picks  $b_j^1$  to minimize

$$\int_{-\pi}^{\pi} |b^0(e^{-i\omega}) - b^1(e^{-i\omega})|^2 g_x(e^{-i\omega}) d\omega.$$

Hints: (a) Write  $y_t$  as

$$y_t = \sum_{j=-\infty}^{\infty} b_j^0 x_{t-j} + \varepsilon_t,$$

$E\varepsilon_t x_{t-j} = 0$  for all  $j$ . Then show that

$$E \left( y_t - \sum_{j=-\infty}^{\infty} b_j^1 x_{t-j} \right)^2 = E(z_t^2)$$

where

$$z_t = \sum_{j=-\infty}^{\infty} (b_j^0 - b_j^1) x_{t-j} + \varepsilon_t.$$

(b) Apply formula (26) to calculate the spectrum of  $z_t$ . (c) Apply formula (20) to calculate the variance of  $z_t$  (see Sims, 1972b).

2. ("Optimal" seasonal adjustment via signal extraction) Suppose that an analyst is interested in estimating  $x_t$  but only observes  $X_t = x_t + u_t$  where  $Ex_t u_s = 0$  for all  $t$  and  $s$ , and where  $x_t$  and  $u_t$  are both covariance stationary stochastic processes with means of zero and known (to the analyst) covariance generating functions  $g_x(z)$  and  $g_u(z)$  respectively;  $g_x(e^{-i\omega}) > 0$  for all  $\omega$ , but has most of its power concentrated at seasonal frequencies. The analyst estimates  $x_t$  by the projection

$$\hat{x}_t = \sum_{j=-\infty}^{\infty} h_j X_{t-j},$$

the projection of the unknown  $x_t$  on the  $X_t$  process.

- A. Derive a formula for the  $h_j$  (use (38)).
- B. Prove that  $g_x(e^{-i\omega}) < g_x(e^{-i\omega})$  for all  $\omega$ .
- C. Prove that if  $g_x(e^{-i\omega})$  is relatively smooth across the seasonal and nonseasonal frequencies, then since  $g_u(e^{-i\omega})$  has big peaks at the seasonal frequencies, it follows that  $g_x(e^{-i\omega})$  will have substantial dips at the seasonal frequencies.

3. Let  $x_t$  be any covariance stationary stochastic process with  $Ex_t = 0$ .

- A. Prove that there exists a representation

$$x_t = \sum_{j=0}^{\infty} c_j u_{t+j} + \theta_t$$

where  $c_0 = 1$ ,  $\sum c_j^2 < \infty$ ,  $Eu_t u_s \geq 0$ ,  $Eu_t u_s = 0$  for  $t \neq s$  and  $E\theta_t u_s = 0$  for all  $t$  and  $s$ ;  $\theta_t$  is a process that can be predicted arbitrarily well by a linear function of only future values of  $x_t$  and  $u_t = x_t - P[x_t | x_{t+1}, x_{t+2}, \dots]$ .

- B. Prove that  $c_j = d_j$  where  $d_j$  is the object in Wold's theorem.
- C. Does  $u_t = \varepsilon_t$  where  $\varepsilon_t$  is the object in Wold's theorem? Does  $Eu_t^2 = E\varepsilon_t^2$ ?
- D. Does  $\theta_t = \eta_t$  where  $\eta_t$  is the object in Wold's theorem?

4. Consider the "explosive" first-order Markov process  $y_t = \lambda y_{t-1} + \varepsilon_t$ ,  $t = 1, 2, \dots$ ,  $\lambda > 1$ , where  $\varepsilon_t$  is white noise with mean zero and variance  $\sigma_\varepsilon^2$ , and  $y_0$  is given.

- A. Prove that for each realization  $(\varepsilon_1, \varepsilon_2, \dots)$  the  $y_t$  process has the representation

$$y_t = \lambda^t \eta_0 + \frac{1}{1-\lambda^{-1}L} u_t$$



where  $u_t$  is a white noise. Find formulas for  $\eta_0$  and  $u_t$  in terms of the  $\varepsilon$  process,  $\lambda$ , and  $y_0$ . (Hint: solve the difference equation forward and impose the initial condition.)

B. Is the  $u_t$  process "fundamental" for  $y_t$ ?

5. Consider the univariate first-order mixed moving average, autoregressive process  $z_t = \lambda z_{t-1} + a_t - \beta a_{t-1}$  where  $a_t$  is a fundamental white noise for  $z$  and  $0 < \beta < 1$ ,  $0 < \lambda < 1$ .

A. Write the process in the form (65). (Hint: try  $x_t = (z_t, a_t)'$  and  $\varepsilon_t = (a_t, a_t)'$ .)

B. Use formula (70) to derive a formula for  $P[z_{t+2}|z_t, z_{t-1}, \dots]$ . Verify that this answer agrees with the result of applying the Wiener-Kolmogorov formula (52).

6. For the processes below, determine whether  $x$  Granger causes  $y$  and whether  $y$  Granger causes  $x$ .

A.

$$g_x(z) = \sigma_\varepsilon^2 \frac{1}{1 - 0.9z} \frac{1}{1 - 0.9z^{-1}}, \quad g_y(z) = \sigma_u^2 (1 - 0.8z)(1 - 0.8z^{-1}),$$

$$g_{yx}(z) = \sigma_{ue}(1 - 0.8z)(1 + 0.5z^{-1}).$$

B.

$$g_x(z) = \sigma_\varepsilon^2 (1 + 0.99z)(1 + 0.99z^{-1}), \quad g_y(z) = \sigma_u^2 \left( \frac{1}{1 - 0.7z + 0.3z^2} \right) \left( \frac{1}{1 - 0.7z^{-1} + 0.3z^{-2}} \right),$$

$$g_{yx}(z) = \sigma_{ue}(1 + 0.2z)(1 + 0.99z^{-1}).$$

C.

$$g_x(z) = \sigma_\varepsilon^2 \left( \frac{1}{1 - 0.7z} \right) \left( \frac{1}{1 - 0.7z^{-1}} \right), \quad g_y(z) = \sigma_u^2 \left( \frac{1}{1 - 0.8z} \right) \left( \frac{1}{1 - 0.8z^{-1}} \right),$$

$$g_{yx}(z) = \sigma_{ue} \left( \frac{1}{1 - 0.8z} \right) \left( \frac{1}{1 - 0.7z^{-1}} \right).$$

7. Consider the simple Keynesian macroeconomic model

$$c_t = \sum_{j=0}^{\infty} b_j Y_{t-j} + \varepsilon_t, \quad \sum_{j=0}^{\infty} b_j^2 < \infty, \quad c_t + I_t = Y_t \quad (*)$$

where  $c_t$ ,  $Y_t$ , and  $I_t$  are consumption, GNP, and investment, respectively, all measured as deviations from their means. Here  $\varepsilon_t$  is a stationary disturbance process that satisfies  $E\varepsilon_t \cdot I_s = E\varepsilon_t = 0$  for all  $t$  and  $s$  and  $I_s$  is a stationary stochastic process. Assume that  $(1 - b(L))$  has a one-sided, square summable inverse in nonnegative powers of  $L$ .

A. Determine whether  $Y$  Granger causes  $I$ .

B. Determine whether  $c$  Granger causes  $Y$  and whether  $Y$  Granger causes  $c$ . (Hint: solve for  $c_t$  and  $Y_t$  each as "reduced form" functions of  $I$  and  $\varepsilon$ , then apply formula (18) to calculate the cross spectrum and use formula (38) to investigate Granger causality.)

C. Is the consumption function (\*) a projection (regression) equation?

8. Consider a  $(y, x)$  process that has a Wold moving average representation

$$y_t = a(L)\varepsilon_t + ka(L)u_t, \quad x_t = c(L)\varepsilon_t$$

where  $k$  is a constant,  $a(L)$  and  $c(L)$  are each one-sided on the past and present and square summable,  $E\varepsilon_t = E\varepsilon_t = E\varepsilon_s = 0$  for all  $t$  and  $s$ , and where  $\varepsilon_t$  and  $u_t$  are jointly fundamental for  $y$  and  $x$ . Finally, assume that both  $a(L)$  and  $c(L)$  are invertible, i.e., have square-summable inverses that are one-sided in nonnegative powers of  $L$ .

A. Determine whether  $y$  Granger causes  $x$  and whether  $x$  Granger causes  $y$ .

B. Find the coefficient generating function for the projection of  $y$  on the entire  $x$  process.

C. Find the coefficient generating function for the projection of  $x$  on the entire  $y$  process.

D. Obtain a different Wold moving average representation for the  $(y, x)$  process. (Hint: choose one white noise process as  $\eta_{1t} \equiv \varepsilon_t + ku_t$ , and choose the other as  $\eta_{2t}$ , the error in the projection of  $\varepsilon_t$  on  $\varepsilon_t + ku_t$ ,  $\varepsilon_t = \rho(ku_t + \varepsilon_t) + \eta_{2t}$ , where  $\eta_{2t}$  is a least squares disturbance.)

9. Consider Lucas's aggregate supply curve

$$y_t = y(p_t - P[p_t|\Omega_{t-1}]) + \lambda y_{t-1} + u_t, \quad 0 < |\lambda| < 1, \quad \gamma > 0 \quad (*)$$

where  $y$  is the log of real GNP,  $p$  the log of the price level, and  $u_t$  a stationary random disturbance process. Suppose that  $p_t$  follows the Markov process

$$p_t = \sum_{i=1}^n w_i p_{t-i} + \varepsilon_t \quad (\dagger)$$

where  $P[\varepsilon_t|\Omega_{t-1}] = 0$ . Here  $\Omega_{t-1}$  is an information set including at least lagged  $y$ 's and lagged  $p$ 's.

A. Suppose that  $P[u_t|\Omega_{t-1}] = 0$ , so that  $u_t$  is serially uncorrelated. Prove that  $p$  fails to Granger cause  $y$ . (In fact, this can be proved where  $p$  follows any arbitrary stationary stochastic process and is not dependent on  $p$  following  $(\dagger)$ .)

B. Now assume  $(\dagger)$  and suppose that  $u_t$  is serially correlated, and in particular that

$$u_t = \rho u_{t-1} + \xi_t, \quad 0 < |\rho| < 1$$

where  $P[\xi_t|\Omega_{t-1}] = 0$ . Prove that  $p$  Granger causes  $y$  by calculating

$$P[y_t|y_{t-1}, y_{t-2}, \dots, p_{t-1}, p_{t-2}, \dots].$$

10. Suppose that  $y_t$  fails to Granger cause  $x_t$ , where both  $y$  and  $x$  are seasonally unadjusted processes. Suppose that an investigator studies seasonally adjusted processes  $y_t^a$  and  $x_t^a$  (see Sims, 1974):

$$y_t^a = f(L)y_t, \quad x_t^a = g(L)x_t$$

where  $f(L)$  and  $g(L)$  are each finite-order two-sided, symmetric ( $f_j = f_{-j}$ ,  $g_j = g_{-j}$ ) seasonal adjustment filters chosen so that  $y_t^a$  and  $x_t^a$  have less power at the seasonal frequencies than do  $y_t$  and  $x_t$ , respectively. Assume that  $y_t^a$  and  $x_t^a$  are strictly linearly indeterministic, as are  $y_t$  and  $x_t$ .

Prove that if  $f(L) \neq g(L)$ , then  $y_t^a$  in general Granger causes  $x_t^a$ . (Hint: first calculate the coefficient generating function for the projection of  $y_t$  on the  $x$  process, then calculate the coefficient generating function for the projection of  $y_t^a$  on the  $x^a$  process.)

11. A recent paper claims that permanent income  $Y_{pt}$  theoretically follows a "random walk," i.e., it obeys

$$P_t Y_{pt+1} = Y_{pt}, \quad (*)$$

where  $P_t Y_{pt+1} \equiv P Y_{pt+1} | \Omega_t$  and  $\Omega_t \supset Y_{pt}$ . Suppose we define  $Y_{pt}$  as

$$Y_{pt} = (1 - \beta) \sum_{j=0}^{\infty} \beta^j P_t Y_{t+j}, \quad 0 < \beta < 1. \quad (\dagger)$$

Assume that  $\Omega_t \supset \Omega_{t-1} \supset \Omega_{t-2} \supset \dots$ .

A. Is  $(*)$  in general an implication of  $(\dagger)$ ? (Hint: apply the law of iterated projections.)

B. Can you restrict the  $\{Y_t\}$  stochastic process so that  $(*)$  is an implication of  $(\dagger)$  in a special case?

12. In a recent article, a macroeconomist reported a regression of the log of the price level ( $p_t$ ) on current and past values of the log of the money supply ( $m_t$ ):

$$p_t = a + \sum_{j=0}^{\infty} h_j m_{t-j} + \varepsilon_t, \quad E\varepsilon_t m_{t-j} = 0 \quad \text{for } j \geq 0 \quad (*)$$

$$E\varepsilon_t = 0$$

where  $e_t$  is a random disturbance. He found that the  $h_j$  were nonzero for many  $j$ 's. He concluded that prices are "too sticky" to be explained by an equilibrium model. According to this economist, "classical" macroeconomics implies that  $h_0 = 1$  and  $h_j = 0$  for  $j \neq 0$ .

Now consider the following classical macroeconomic model:

$$m_t - p_t = \alpha(P_t p_{t+1} - p_t) + y_t + u_t \quad (\text{portfolio balance schedule})$$

$y_t = \text{constant}$  (extreme classical full-employment assumption). Here  $\alpha < 0$ , and  $u_t$  is a stationary random process obeying

$$Eu_t m_t = 0 \quad \text{for all } t, s, \quad Eu_t = 0.$$

The money supply is exogenous and has moving average representation

$$m_t = d(L)e_t, \quad e_t = m_t - Pm_t | m_{t-1}, m_{t-2}, \dots, \quad \text{and} \quad \sum_{j=0}^{\infty} d_j^2 < \infty.$$

Derive a formula giving the  $h(L) = \sum_{j=0}^{\infty} h_j L^j$  in (\*) as a function of  $\alpha$  and  $d(L)$ . Is the macro-economist correct in his interpretation of the implications of classical theory?

13. Let the portfolio balance schedule be Cagan's

$$\mu_t - x_t = \alpha(P_t x_{t+1} - P_{t-1} x_t) + \eta_t \quad (*)$$

where  $\mu_t$  is the rate of growth of the money supply,  $x_t$  is the rate of inflation, and  $\eta_t$  satisfies  $P_{t-1}\eta_t = 0$ , where  $P_t[y] \equiv P[y | \mu_t, \mu_{t-1}, \dots, x_t, x_{t-1}, \dots]$  in which  $y$  is any random variable. (Equation (\*) is just the first difference of Equation (57) in the text.)

A. Prove that a solution of (\*) is

$$P_t x_{t+1} = \frac{1}{1-\alpha} \sum_{j=1}^{\infty} \left( \frac{-\alpha}{1-\alpha} \right)^{j-1} P_t \mu_{t+j}. \quad (\dagger)$$

B. Suppose that  $(x_t, \mu_t)$  has the bivariate vector moving average, autoregressive representation

$$\begin{pmatrix} x_t \\ \mu_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1-\lambda & \lambda \end{pmatrix} \begin{pmatrix} x_{t-1} \\ \mu_{t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} - \lambda a_{1t-1} \\ a_{2t} - \lambda a_{2t-1} \end{pmatrix}$$

where  $a_{1t} = x_t - P_{t-1}x_t$ ,  $a_{2t} = \mu_t - P_{t-1}\mu_t$ ,  $|\lambda| < 1$ , and  $a_{1t}$  and  $a_{2t}$  have finite variances and non-zero covariance. Prove that Cagan's formula for the expected rate of inflation  $\pi_t$ ,

$$\pi_t = \frac{1-\lambda}{1-\lambda L} x_t$$

is implied by the hypothesis of rational expectations, i.e., by Equation (†).

C. Prove that  $\mu$  fails to Granger cause  $x$ .

D. Calculate the coefficients in the projection of  $\mu_t - x_t$  on the  $x_t$  process. Is this projection equation the same as Cagan's equation

$$\mu_t - x_t = [\alpha(1-\lambda)/(1-\lambda L)](1-L)x_t + \xi_t, \quad (\S)$$

where  $\xi_t$  is random? If not, use your formula for the projection equation to determine the biases that would emerge from mistakenly regarding Cagan's (§) as a projection equation.

## REFERENCES

- Allen, R. G. D. (1960). *Mathematical Economics*, 2nd ed., New York: Macmillan.  
 Andersen, L. C., and Jordan, J. L. (1968). "Monetary and fiscal actions: A test of their relative importance in economic stabilization." *Federal Reserve Bank of St. Louis Review*, Vol. 50, No. 11, pp. 11-23.

- Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, New York: Wiley.  
 Apostol, T. M. (1974). *Mathematical Analysis*, 2nd ed., Reading, Massachusetts: Addison-Wesley.  
 Cagan, P. (1956). "The monetary dynamics of hyperinflation." *Studies in the Quantity Theory of Money* (M. Friedman, ed.), Chicago, Illinois: University of Chicago Press.  
 Chow, G. C., and Levitan, R. E. (1969). "Nature of business cycles implicit in a linear economic model." *The Quarterly Journal of Economics*, Vol. LXXXIII, No. 3, pp. 504-517.  
 Frisch, R. (1933). *Economic Essays in Honor of Gustav Cassel*, London: Allen and Unwin.  
 Granger, C. W. J. (1966). "The typical spectral shape of an economic variable." *Econometrica*, Vol. 34, No. 1, pp. 150-161.  
 Granger, C. W. J. (1969). "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica*, Vol. 37, No. 3, pp. 424-438.  
 Howrey, E. P. (1968). "A spectrum analysis of the long-swing hypothesis." *International Economic Review*, Vol. 9, No. 2, pp. 228-260.  
 Howrey, E. P. (1971). "Stochastic properties of the Klein-Goldberger model." *Econometrica*, Vol. 39, No. 1, pp. 73-88.  
 Muth, J. F. (1960). "Optimal properties of exponentially weighted forecasts." *Journal of the American Statistical Association*, Vol. 55, No. 290, pp. 299-306.  
 Naylor, A. W., and Sell, G. R. (1971). *Linear Operator Theory in Engineering and Science*, New York: Holt.  
 Shiller, R. (1972). "Rational Expectations and the Structure of Interest Rates," unpublished Ph.D. Thesis, Cambridge, Massachusetts: M.I.T.  
 Sims, C. A. (1972a). "Money, income, and causality." *The American Economic Review*, Vol. LXII, No. 4, pp. 540-552.  
 Sims, C. A. (1972b). "Approximate prior restrictions in distributed lag estimation." *Journal of the American Statistical Association*, Vol. 67, No. 337, pp. 169-175.  
 Sims, C. A. "Seasonality in regression." *Journal of the American Statistical Association*, Vol. 69, No. 347, pp. 618-626.  
 Slutsky, E. (1937). "The summation of random causes as the source of cyclic processes." *Econometrica*, Vol. 5, pp. 105-146.  
 Theil, H. (1971). *Principles of Econometrics*, pp. 548-550, New York: Wiley.  
 Whittle, P. (1963). *Prediction and Regulation by Linear Least-Square Methods*, Princeton, New Jersey: Van Nostrand-Reinhold.  
 Wold, H. (1938). *The Analysis of Stationary Time Series*, 1st ed., Uppsala: Almqvist and Wicksell.