Federal Reserve Bank of Minneapolis
Research Department Working Paper

NOTES ON CONTROL AND PREDICTION

Thomas J. Sargent

Working Paper 229
PACS File 2700

November 1982

9. A More General Univariate Optimization Problem

The problem of the preceding section is a special case of a more general quadratic optimization problem, one version of which follows. We define the polynomial in the lag operator,

(58) $$d(L) = d_0 + d_1 L + \cdot \cdot \cdot + d_m L^m,$$

where $d_0 \neq 0$, $d_m \neq 0$. We assume that g_t is a sequence of exponential order less than $1/\sqrt{b}$, where $0 < b < 1$. Then the problem is to choose a sequence $\{y_t, t \geq 0\}$ to maximize

(59) $$\sum_{t=0}^{\infty} b_t \{g_t y_t - \frac{1}{2} h y_t^2 - \frac{1}{2} [d(L) y_t]\}$$

where $h > 0$,

subject to $y_{-1}, y_{-2}, \cdot \cdot \cdot y_{-m}$ given.

The problem of Section 8 is a version of problem (59) with $g_t = (f_0 + a_t - w_t)$, $m = 1$, $y_t = n_t$, and $d(L) = \frac{d}{2}(1-L)$, except that in the problem of this section we have strengthened the requirement on g_t that it be of exponential order less than $1/\sqrt{b}$, rather than $1/b$ (note that $1/\sqrt{b} < 1/b$). The reason that this stronger condition is needed for the more general problem of this section will be described below.

The Euler equation for this problem is

(60) $$\left[d(bL^{-1})d(L) + h\right] y_t = g_t.$$

We invite the reader to verify that this is the Euler equation by differentiating the right side (59) with respect to y_t and rearranging. Note that (53) is a special case of (60) with

$d(L) = (1-L)\sqrt{d/2}$, $h = f_1$, $gt = (f_0 + at - wt)$.

In addition to the Euler equation (60), the necessary and sufficient conditions for maximization of (61) are completed by the condition

$$(61) \qquad \frac{h}{2} \sum_{t=0}^{\infty} b^t y_t^2 < +\infty.$$

Inspection of (59) reveals that any $\{y_t\}$ path that violates (61), even if it satisfies the Euler equation, gives a very bad outcome for the criterion function. In general, condition (61) is <u>not</u> implied by the transversality conditions, which are obtained as in Section 8, by differentiating a finite T version of (59) with respect to the $y_T, \ldots y_{T-m+1}$, setting the results to zero, and taking limits as $T \to \infty$. Hansen and Sargent (   ) describe in detail why the transversality conditions, though necessary for an optimum, are not in general sufficient. They also describe how it is that (61) is the general condition that selects the unique solution of the Euler equations (60) that maximizes (59).

We briefly describe how (60) is solved subject to (61). We first note a simple but important feature of the characteristic polynomial $h + d(bL^{-1})d(L)$ that appears in (60). This polynomial evaluated at any value z_0 equals the polynomial evaluated at bz_0^{-1}. In particular, if z_0 is a zero of this polynomial, then so is bz_0^{-1}. To prove this, suppose that z_0 is a zero, which means setting $L = z_0$ gives $h + d\,(bz_0^{-1})\,d(z_0) = 0$. The claim is that this implies that setting $L = bz_0^{-1}$ will also set the polynomial to zero. Making this substitution gives $h + d\,(z_0)$ $d(bz_0^{-1})$, which equals the characteristic polynomial evaluated at z_0, which equals zero by assumption. Thus, the zeros of the characteristic polynomial of the Euler equation come in pairs of the form z_k, bz_k^{-1}, $k = 1, \ldots, m$.

Let us assume that the zeros of the characteristic polynomial are distinct. (The reader can apply the methods of Section 5 to analyze the case of repeated zeroes, if this becomes necessary.) Let the zeroes be ordered in descending order in absolute value by

$|z_1| > |z_2| > \quad > |z_m| > |bz_m^{-1}| > \quad > |bz_2^{-1}| > |bz_1^{-1}|$. Let us define $\lambda_k = 1/z_k$, $k = 1, \ldots, m$. Notice that the above ordering, and in particular $|z_m| > |bz_m^{-1}|$, implies that $|\lambda_k| < 1/\sqrt{b}$ for $k = 1, \ldots, m$. Then it is possible to factor the characteristic polynomial of the Euler equation as

$$(62) \qquad h + d(bL^{-1})d(L) = c(bL^{-1})c(L)$$

where

$$(63) \qquad c(L) \equiv c_0(1-\lambda_1 L)(1-\lambda_2 L) \ldots (1-\lambda_m L)$$

$$(64) \qquad c_0 = [(-1)^m \lambda_0/(\lambda_1 \lambda_2 \ldots \lambda_m]$$

and where λ_0 is a constant that is uniquely determined by h and d(L).

Since $|\lambda_k| < 1/\sqrt{b}$ for $k = 1, \ldots, m$, (62) - (63) imply that $h + d(bL^{-1})d(L)$ has been factored into $c(bL^{-1})\ c(L)$, where the zeroes of $c(z)$ exceed $\sqrt{b}$ in modulus, (from (63) the zeroes of $c(z)$ are the $1/\lambda_k$'s) and where the zeroes of $c(bz^{-1})$ are less than $\sqrt{b}$ in modulus.

Using factorization (62), the Euler equation (60) can be represented as

(65) $$c(bL^{-1})c(L)y_t = g_t$$

The unique solution of the Euler equation that satisfies (61) is given by

(66) $$c(L)y_t = c(bL^{-1})^{-1}g_t.$$

or

(66) $$(1-\lambda_1 L), \ldots (1-\lambda_m L)y_t = c_0^{-2}\,[(1-\lambda_1 bL^{-1}) \ldots (1-\lambda_m L^{-1})]^{-1}g_t.$$

A partial fractions representation of $c_0 c(bL^{-1})^{-1}$ is given by

(67) $$[(1-\lambda_1 bL^{-1}) \ldots (1-\lambda_m bL^{-1})] = \sum_{k=1}^{m} \frac{A_k}{(1-\lambda_k bL^{-1})}$$

where

(68) $$A_k = \lim_{z \to \lambda_k b} (1-\lambda_k bz^{-1})\; c(bz^{-1})/c_0.$$

Using (67), solution (66) can be represented as

(69) $$(1-\lambda_1 L) \ldots (1-\lambda_m L)y_t = c_0^{-2} \sum_{k=1}^{m} \left(\frac{A_k}{1-\lambda_k bL^{-1}}\right)^2 g_t.$$

or

(70) $$(1-\lambda_1 L) \ldots (1-\lambda_m L)y_t = c_0^{-2} \sum_{k=1}^{m} A_k \sum_{j=0}^{\infty} (b\lambda_k)^j g_{t+j}.$$

Since $|\lambda_k| < 1/\sqrt{b}$, $|b\lambda_k| < \sqrt{b}.$ This condition, together with our having assumed that $\{g_t\}$ is of exponential order less than $1/\sqrt{b}$ guarantees that the right hand side converges, and that as a function of t it is itself of exponential order less than $1/\sqrt{b}.$ This condition, together with the fact that

$|\lambda_k| < 1/\sqrt{b}$, $k = 1, \ldots, m$, guarantees that the solution (70) starting from the given starting values $y_{-1}, \ldots, y_{-m}$ satifies condition (61) and is therefore optimal.

We now briefly describe why we were able to get by with the weaker assumption of exponential order less than 1/b, rather than $1/\sqrt{b}$, for $\{a_t\}$ and $\{w_t\}$ in the problem of Section 8. The reason is simply that for the special polynomial $d(L) = \sqrt{d/2}\,(1-L)$, it turns out that for any $h > 0$, the factorization (62) - (63) is such that $|\lambda_1| < 1$, which is stronger than the general condition that $|\lambda_1| < 1/\sqrt{b}$. The reader can verify that with more general polynomials of the form $d(L) = \sqrt{d/2}\,(1-\alpha L)$, values of α exist for which the analogue of $-\phi$ in figure 4 is less than $(1+b)$, so that $1 < \lambda_1 < 1/\sqrt{b}$. For such cases, it is necessary to impose a stronger condition on the exponential orders of (a_t, w_t) than we did in Section 8.

We close this section with a remark on terminology. Consider solution (66), $c(L)y_t = c(bL^{-1})^{-1}g_t$. In the control literature, $c(L)y_t$ is called the "feedback part" of the solution for y, while $c(bL^{-1})^{-1}g_t$ is called the "feed forward part."

10. Introduction to Multivariate Dynamic Optimization

We now briefly describe a multivariate generalization of the problem of Section 9, and its solution. We define the matrix polynomial in the lag operator

$$D(L) = D_0 + D_1 L + \ldots + D_m L^m \tag{71}$$

where D_j is an (nxn) matrix. We let $\{G_t\}$ be a sequence of (mx1) vectors, each component of which is a sequence of exponential order less than $1/\sqrt{b}$. We let H be an (nxn) positive definite matrix. Finally, we let $\{Y_t\}$ be a sequence of (mx1) vectors of variables that are to be chosen for $t > 0$, with given initial values $Y_{-1}, Y_{-2}, \ldots, Y_{-m}$. The problem we are interested in is to choose $\{Y_t, t \geq 0\}$ to maximize

$$(72) \qquad \sum_{t=0}^{\infty} b^t \{G_t' Y_t - \frac{1}{2} Y_t' HY_t - \frac{1}{2} [D(L)Y_t]' [D(L)Y_t]\}$$

given $\{G_t, t \geq 0\}$ and $Y_{-1}, \ldots, Y_{-m}$. In (72), the prime (′) denotes matrix transposition.

Necessary and sufficient conditions for a maximum of (72) are

$$(73) \qquad \sum_{t=0}^{\infty} b^t Y_t' HY_t < +\infty$$

and the Euler equations

$$(74) \qquad \{H + D(bL^{-1})' D(L)\} Y_t = G_t$$

That the Euler equations assume the form (74) can be proved using the method of Sections 8 and 9, namely by differentiating (72) with respect to Y_t, equating the result to a vector of zeroes, and rearranging. Condition (73) is the correct generalization of (61), and is justified by similar reasoning.

The related pairs property of the zeroes of the characteristic polynomial of the Euler equation that held in the univariate case generalizes as folllows. If z_0 is a zero of det $\{H + D(bz^{-1})' D(z)\}$ then so is bz_0^{-1}. Here "det" denotes the determinant of a matrix. The appropriate matrix analogue of the scalar polynomial factorization (62) is a polynomial matrix factorization $C(bL^{-1})' C(L)$, where the zeroes of det $C(z)$ exceed $\sqrt{b}$, in absolute value, while these of det $C(bz^{-1})$ are less than $\sqrt{b}$ in absolute value. By a theorem on matrix polynomials of the form that appear in (74), there always exists a matrix factorization

(75) $$H + D(bL^{-1})' D(L) = C(bL^{-1})' C(L)$$

where

(76) $$C(L) = C_0 + C_1 L + \ldots + C_{\bar{m}} L^{\bar{m}}$$

and where the zeroes of det $C(z)$ exceed $\sqrt{b}$.

The solution of (74) satisfying (73) is then given by

(77) $$C(L)Y_t = C(bL^{-1})'^{-1} G_t.$$

By the above mentioned property about the location of the zeroes of det $C(bz^{-1})$, we can represent det $C(bL^{-1})'$ as

$$\det C(bL^{-1})' = \lambda_0 (1-\lambda_1 bL^{-1}) \ldots (1-\lambda_k bL^{-1})$$

where $|\lambda_1| < 1/\sqrt{b}$, where $k = m\ n$, and where we have assumed distinct zeroes of det $C(bz^{-1})$. Using a matrix version of partial fractions, $C(bL^{-1})'^{-1}$ can be represented as

(78) $$C(bL^{-1})'^{-1} = \sum_{h=1}^{k} \frac{A_h}{(1-\lambda_h bL^{-1})}$$

where

(79) $$A_h = \lim_{z \to \lambda_h b} (1-\lambda_h bz^{-1})\ C(bz^{-1})'^{-1}$$

Note that each A_h is an (nxn) matrix. Using (78), (79) can be represented

$$C(L)Y_t = \sum_{h=1}^{\infty} A_h \sum_{j=0}^{\infty} (\lambda_n b)^j G_{t+j} \tag{80}$$

Representation (80) is the vector generalization of (70).

Note in (77) a sort of symmetry in form between the "feedback part" $C(L)Y_t$ and the "feed forward part" $C(bL^{-1})'^{-1}G_t$ that generalizes a similar symmetry that we observed in the univariate problems.

A key step in solving problems of this sort is achieving the matrix factorization (75). Hansen and Sargent [ ] describe several methods for accomplishing this. The most readily understandable one is probably the one that uses iterations on a "matrix Riccati difference equation" (see Chapter ______, pp.). For now it is sufficient for the reader to trust that practical methods exist for factoring $H + D(bL^{-1})'D(L)$ in the manner required.

Example (i) Interrelated Factor Demand

Consider the problem of a firm that maximizes

$$\sum_{t=0}^{\infty} b^t\{q_t - w_t n_t - J_t K_t - \frac{d_1}{2}[(1-L)n_t]^2 - \frac{d_2}{2}[(1-L)k_t]^2\} \qquad 0 < b < 1 \tag{81}$$

subject to

$$q_t = f_1' \begin{pmatrix} n_t \\ k_t \end{pmatrix} - \frac{1}{2} \begin{pmatrix} n_t \\ k_t \end{pmatrix}' F \begin{pmatrix} n_t \\ k_t \end{pmatrix}. \tag{82}$$

where k_t is the stock of capital, n_t is the stock of labor, w_t is the real rental on labor, and J_t is the real rental on capital; f_1 is a (2x1) vector of positive constants, F is a positive definite matrix, and d_1 and d_2 are positive constants measuring adjustment costs. We assume that $\{J_t\}$ and $\{w_t\}$ for $t \geq 0$ are known sequences of exponential order less than $1/\sqrt{b}$. The problem is to choose sequences $\{k_t, n_t, t \geq 0\}$ to maximize (81) subject to (82), given initial values n_{-1}, k_{-1} and given sequences for w_t and J_t.

Problem (81) - (82) is a special case of (72) with

$$G_t = f_1 - \begin{pmatrix} w_t \\ n_t \end{pmatrix}, \quad D(L) = \begin{pmatrix} \sqrt{\frac{d_1}{2}}(1-L) & 0 \\ 0 & \sqrt{\frac{d_2}{2}}(1-L) \end{pmatrix}, \text{ and}$$

H = F. The solution (80) is an interrelated pair of decision rules for (n_t, k_t) of the form

$$(83) \qquad C(L) \begin{pmatrix} n_t \\ k_t \end{pmatrix} = \sum_{k=1}^{4} A_h \sum_{j=0}^{\infty} (\lambda_h b)^j \left[f_1 - \begin{pmatrix} w_{t+j} \\ J_{t+j} \end{pmatrix} \right] .$$

where $C(L) = C_0 + C_1 L$. In (83) the decision rules for capital and labor interact in the sense of each of (n_t, k_t) depending on lagged values of the other, and each depending on future rental rates for the other. This interdependence occurs so long as either F(or H) or D(L) is not diagonal.

Versions of this problem were studied and utilized by Nadiri and Rosen [ ], Hansen and Sargent [ ], Meese [ ], and Eichenbaum [ ].

Example (ii): A Dynamic Nash Equilibrium

We consider a duopoly in which demand is governed by a linear demand schedule.

$$p_t = A_0 - \frac{A_1}{2}(q_{1t} + q_{2t}) + u_t, \; A_0, A_1 > 0 \tag{84}$$

where q_{it} is output of firm i at t, and u_t is a known sequence of disturbances to demand of exponential order less than $1/\sqrt{b}$. Firm i maximizes

$$\sum_{t=0}^{\infty} b^t \{p_t q_{it} - q_{it} s_{it} - [d_i(L)q_{it}]^2\}, \quad 0 < b < 1 \tag{85}$$

where $d_i(L) = d_{i0} + \ldots + d_{im} L^m$, and where s_{it} is a sequence of shocks to costs of production of firm i, assumed to be a known sequence of exponential order less than $1/\sqrt{b}$. Here $[d_i(L)q_{it}]^2$ stands for costs of adjusting production rapidly. The maximization of (85) by i takes $\{s_{it}\}$, $\{u_t\}$, and $\{q_{jt}, j \neq i,\}$ given for $t \geq 0$, and q_{it} given for $\{t = -m, \ldots, -1\}$. In particular, firm i is imagined to regard firm j's output sequence as given and beyond its influence.

Substituting (84) into (85) gives

$$\sum_{t=0}^{\infty} \{A_0 - \frac{A_1}{2}(q_{1t} + q_{2t}) + u_t\} q_{it} - q_{it} s_{it} - [d_i(L)q_{it}]^2\}. \tag{86}$$

The Euler equations for this problem for firms 1 and 2 are

$$\begin{bmatrix} (A_1 + d_1(bL^{-1})d_1(L)) & \frac{A_1}{2} \\ \frac{A_1}{2} & (A_1 + d_2(bL^{-1})d_2(L)) \end{bmatrix} \begin{bmatrix} q_{1t} \\ q_{2t} \end{bmatrix} = \begin{pmatrix} A_0 + u_t - s_{1t} \\ A_0 + u_t - s_{2t} \end{pmatrix} \tag{87}$$

We define a Nash equilibrium in the space of sequences of quantities q_{1t}, q_{2t} as a pair of sequences for q_{1t}, q_{2t} */ that solve the interrelated Euler equations (87) and satisfy the boundary conditions

(88) $\sum b^t q_{it}^2 < +\infty$ for i = 1, 2.

Equivalently, the Nash equilibrium is the pair of (q_{1t}, q_{2t}) sequences that satisfies the following conditions:

(i) Firm i's quantity sequence maximizes its present value (85), given firm j's sequence, for (i,j) = 1,2) and (2,1).

(ii) The output market clears, in the sense that (84) holds for all t.

Equation (87) is evidently in the form of a vector Euler equation in (q_{1t}, q_{2t}). The matrix polynomial on the left side of (87) can be factored into the form $C(bL^{-1})' C(L)$, where $C(L)$ is a (2x2) matrix polynomial with the zeroes of det $C(z)$ exceeding $\sqrt{b}$ in absolute value. Then the Nash equilibrium can be represented

$$C(L) \begin{matrix} q_{1t} \\ q_{2t} \end{matrix} = C(bL^{-1})'^{-1} \begin{matrix} A_0 + u_t - s_{1t} \\ A_0 + u_t - s_{2t} \end{matrix}$$

where

$$C(L) = C_0 + C_1 L + \ldots + C_m L^m.$$

---

*/It is necessary to add the qualifier "in the space of sequences of quantities q1t, q2t" because different definitions of strategy spaces in general give rise to distinct Nash equilibria.

Note that the vector Euler equation (87) was discovered by our having solved an interrelated pair of univariate dynamic optimization problems. The resulting system of Euler equations (87) seemingly can itself be viewed as solving some vector optimization problem, since the matrix characteristic polynomial is expressible in the form $H + D(bL^{-1})' D(L)$. To seek the vector dynamic optimization problem that is implicitly solved by (87) is to pose a version of an "inverse optimal control problem:" given a system of difference equations, attempt to synthesize an optimum problem for which they are necessary conditions.

Another Example: Learning by Doing

The following example is interesting from both the substantive and a technical points of view. From the technical point of view, the example will introduce a control problem in which it is a rational characteristic polynomial for an Euler equation that must be factored.

We consider a monopolist who faces the linear law of demand

$$\text{(a)} \qquad p_t = A_0 - A_1 Q_t + u_t, \; A_0, \; A_1 > 0$$

where p_t is price, Q_t is output, and u_t is a known sequence of shocks to demand assumed to be of exponential order less than $1/\sqrt{b}$. The monopolist's total costs at t are given by

$$\text{(b)} \qquad C(Q_t) = c_0 + c_1 Q_t + \frac{c_2}{2} Q_t^2 + c_3 Q_t s_t - c_4 [h(L) Q_t] Q_t,$$

$$c_0, \; c_1, \; c_2, \; c_3, \; c_4 > 0$$

where s_t is a known sequence of shocks to marginal cost, assumed to be of exponential order less than $1/\sqrt{b}$, where $h(L) = \sum_{j=0}^{\infty} h_j L^j$. As an example of the sort of $h(L)$ that we have in mind, we shall later consider the special case $h(L) = 1/(1-\rho L)$ where $\rho < 1$ but where ρ is close to one. Then (6) captures the notion that marginal costs of current output fall with cumulated past output. This is one version of a "learning-by-doing" cost structure.

The firm maximizes

$$\sum_{t=0}^{\infty} b^t \{p_t Q_t - C(Q_t)\}$$

or using (a) and (b)

(c) $$\sum_{t=0}^{\infty} b^t\{[A_0 - \frac{A_1}{2} Q_t + u_t]Q_t - [c_0 + c_1 Q_t + \frac{c_2}{2} Q_t^2 + c_3 Q_t s_t - c_4[h(L)Q_t]Q_t]\}.$$

By using the preceding method, the Euler equation for the firm's problem is found to be, after some rearrangement,

(d) $$[(A_1+c_2) - c_4 h(L) - c_4 h(bL^{-1})]Q_t = A_0 + u_t - c_3 s_t.$$

The boundary condition is

(e) $$A_1 \sum_{t=0}^{\infty} b^t Q_t^2 < +\infty.$$

We shall now consider the special case of the model that emerges when we set $h(L) = 1/(1-\rho L)$, $0 < \rho < 1$. In this case, the characteristic polynomial on the left side of the Euler equation (d) become the rational polynomial

$$k - \frac{c_4}{1-\rho L} - \frac{c_4}{1-\rho bL^{-1}}$$

where $k \equiv (A_1+c_2)$. To solve the Euler equation (d) subject to boundary condition (e), the first step is to express the characteristic polynomial in terms of a common denominator, which gives

(f) $$\frac{[k(1+\rho^2\beta)-2c_4] - (k-c_4)L - (k-c_4)bL^{-1}}{(1-\rho bL^{-1})(1-\rho L)}.$$

Notice that the denominator is already factored, and that the zeroes of the numerator come in the familiar type of pairs (z_0, bz_0^{-1}). Our next step on the way to solving our problem is to factor the numerator. Note that the numerator can be written

$$\text{(g)} \qquad (k-c_4)\rho b[-L^{-1}+[\frac{k(1+\rho^2 b)-2c_4}{(k-c_4)\rho b}]-1/bL] \equiv \frac{1}{\alpha b}(1-\alpha bL^{-1})(1-\alpha L)$$

where $1/\alpha$ is the zero of the characteristic polynomial that exceeds $1/\sqrt{b}$. Equating powers of L, as in section 8, shows that α must solve

$$\text{(h)} \qquad \frac{1}{\alpha\beta} + \alpha = [\frac{k(1+\rho^2 b)-2c_4)}{(k-c_4)\rho b}].$$

We assume that the parameter $k \equiv A_1 + c_2$, c_4, ρ, b are such that the right side exceeds $2\sqrt{b}$ in absolute value. This guarantees the existence of a real value α that solves (h). Note that for $\rho = b = 1$, the above equation has the solution $\alpha = 1/\alpha = 1$. By continuity of the solution in the arguments on the right hand side, for values of ρ and b sufficiently close to 1, α will be close to one.

Using (f) and (g), the Euler equation can be expressed

$$\text{(j)} \qquad \frac{(1-\alpha bL^{-1})(1-\alpha L)}{(1-\rho bL^{-1})(1-\rho L)} Q_t = \frac{\alpha b}{(k-c_4)\rho b} [A_0+u_t-c_3 s_t]$$

The solution of the Euler equation (d) that satisfies boundary condition (e) can be expressed in "feedback-feedforward" form

$$\text{(k)} \qquad \frac{(1-\alpha L)}{(1-\rho L)} Q_t = \frac{\alpha b}{(A_1+c_2-c_4)\rho_b} \cdot \frac{(1-\rho bL^{-1})}{(1-\alpha bL^{-1})} [A_0+u_t-c_3 s_t].$$

Unless $\alpha = \rho$, Q_t feeds back on an infinite number of its own past values, reflecting the dynamics of the firm's optimally coping with the learning-by-doing cost structure. In general, $\rho \neq \alpha$. However, in the special limiting case $\rho = b = 1$, it can be verified that $\rho = \alpha$. In this special case, (k) collapses to the static decision rule

$$Q_t = \frac{b}{A_1 + c_2 - c_4} [A_0 + u_t - c_3 s_t],$$

despite the presence of the learning-by-doing cost structure.

A version of figure 4 with b = 1 determines the intriguing golden ratio or golden section. The golden ratio is the unique positive number λ whose reciprocal equals itself plus unity: $\lambda^{-1} = 1 + \lambda$. This equation can be rearranged to read $\lambda^{-1} + \lambda = 1 + 2\lambda$. From the quadratic formula, the golden ratio equals $\frac{\sqrt{5} - 1}{2}$ = and is found as the intersection in the positive quadrant of the line $1 + 2\lambda$ with the curve $\lambda + \lambda^{-1}$. The golden ratio, which appears repeatedly in nature and mathematics, fascinated the ancients, and is said to be reflected in the design of the Parthenon. One place that the number occurs in mathematics is as the limit as t goes to infinity of the ratio of successive terms in a Fibonacci sequence. A Fibonacci sequence is generated by the difference equation $x_{t+1} = x_t + x_{t-1}$ with initial conditions $x_0 = 1$, $x_{-1} = 0$. The characteristic polynomial $(1 - L - L^2)$ associated with this equation can be factored as $(\lambda - L)(\lambda^{-1} + L)$ where λ is the golden ratio. For more on the golden ratio, see "Math and Music: The Deeper Links", New York Times, Sunday, Aug. 29, 1982.

More Exercises

1. (Advertising)

1. A monopolist faces the following demand curve for his product,

$$p_t = A_0 - A_1 Q_t + g(L)a_t + u_t; \; A_0, \; A_1 > 0$$

where p_t is price, Q_t is output, q_t is advertising, u_t is a sequence of shocks to demand, and $g(L) = g_0 + g_1 L + \dots + g_m L^m$, where $g_j > 0$ for $j = 0, \dots, m$. The firm maximizes

$$(1) \qquad \sum_{t=0}^{\infty} \beta^t \{p_t Q_t - Q_t s_t - 1/2[d(L)Q_t]^2 - \frac{\gamma}{2} a_t^2 - a_t w_t\}, \; 0 < \beta < 1$$

where $d(L) = \sum_{j=0}^{n} d_j L^j$. In (1), s_t is a shock to costs, $1/2[d(L)Q_t]^2$ represents costs of rapid adjustment, and the marginal costs of advertising at t are $(w_t + \gamma a_t)$, where w_t is a known sequence. We assume that (u_t, s_t, w_t) are known sequences of exponential order less than $1/\sqrt{\beta}$. The criterion (2) is to be maximized over sequences for $\{Q_s, a_s, s \geq 0\}$ taking as given $\{Q_s, a_s, s<0\}$.

a. Find the Euler equations for this problem.

b. Argue that the solution of this will be linear laws of motion for (Q_t, a_t) in which each of (Q_t, a_t) depend on lagged values of both Q and a, and current and future values of all of (u,s,w).

2. (Time to build with two processes)

Consider a monopolist whose output satisfies

$$(1) \qquad Q_t = f(L)n_{1t} + g(L)n_{2t}$$

where $f(L) = \sum_{j=0}^{m} f_j L^j$, $g(L) = \sum_{j=0}^{r} g_j L^j$; $f_j \geq 0$, $g_j \geq 0$ for all j.

In (1), n_{1t} is the amount of labor at k that is assigned to process 1, while n_{2t} is the amount that is assigned to process 2. The idea is that output can be produced via two processes, with different timing characteristics, e.g., to represent the notion that the first process is fast but wasteful, while the other is efficient but time consuming, we might set $f(L) = L$, $g(L) = 1/2[L+L^2+L^3+L^4]$. The firm faces the demand curve

$$(2) \qquad p_t = A_0 - A_1 Q_t + u_t, \ A_0, \ A_1 > 0$$

where u_t is a known sequence of exponential order less than $1/\sqrt{\beta}$. The firm hires labor at the wage rate w_t, where w_t is a known sequence of exponential order less than $1/\sqrt{\beta}$. The firm's problem is to maximize

$$(3) \qquad \sum_{t=0}^{\infty} \beta^t \{p_t Q_t - w_t Q_t\}, \ 0 < \beta < 1$$

subject to (1) and (2), with $\{n_{1-s}, s=-1,\ldots,-m\}$ $\{n_{2s}, s=-1,\ldots,-r\}$ given.

a. Find the Euler equations for this problem.

b. Indicate the form of the optimum decision rules for (n_{1t}, n_{2t}).

Insert A

Among those solutions for $\{n_t\}$ that satisfy the Euler equations, it turns out that the conditions that $\{a_t\}$, $\{w_t\}$, $\{n_t\}$ are also necessary for the transversality condition (54) to hold.

Insert B

The first-order conditions (51)-(52) are necessary and sufficient conditions for maximizing the finite T criteria (50). (The reader can verify that the second order conditions for a maximum are satisfied.) For this particular problem, although not in general, it turns out that the limits of the first-order conditions (51)-(52) as T approaches infinity are necessary and sufficient conditions for a maximum of the infinite horizon problem (49). In more general linear-quadratic optimization problems, the limits as $T \to \infty$ of the first-order conditions are necessary conditions for maximization of the infinite time problem, but are not in themselves sufficient. Hansen and Sargent [ ] discuss this point and provide examples. The special features of the present problem that makes the limits as $T \to \infty$ of the first-order conditions necessary and sufficient for maximizing the infinite T problem are (i) the condition that $f_1 > 0$, and (ii) the fact the zero of the characteristic polynomial of the first difference polynomial that governs adjustment costs, being exactly unity, is not less than unity in absolute value. By violating condition (ii), Hansen and Sargent [ ] provide an example in which the transversality condition and Euler equation fail to be sufficient conditions for an optimum.

We now turn to a method for solving the Euler difference equation subject to the initial condition and the transversality equation.

Finding a Wold Representation: m^{th} order moving average

More generally, suppose that we are given a process u_t with finite order moving average representation

$$x_t = a(L)\, u_t$$

where

$$a(L) = \sum_{j=0}^{m} a_j\, L^j, ,$$ and where u_t is a white noise that is not fundamental for x_t, i.e., $u_t \neq x_t - P[x_t | x_{t-1} \ldots]$. For convenience, we assume that the zeroes of the polynomial $a(z)$ are distinct. The covariance generating function of x_t is given by

(1) $$g(z) = a(z)a(z^{-1})\sigma_u^2.$$

We know that this process also possesses a Wold moving average representation

(2) $$x_t = d(L)\varepsilon_t, \; d_0 = 1,$$

where $d_0\, \varepsilon_t = x_t - P[x_t | x_{t-1}, \ldots]$. The condition that ε_t lie in the linear space spanned by $\{x_t, x_{t-1}, \ldots\}$ is equivalent with the condition that the zeroes of $d(z)$ not exceed unity in absolute value. To see this heuristically, represent $d(L)$ as

(3) $$d(L) = \sigma_\varepsilon(1-\lambda_1 L) \ldots (1-\lambda_m L)$$

where λ_j is the reciprocal of the m^{th} zero of $d(z)$. Represent $d(L)^{-1}$ as

$$(4) \qquad d(L)^{-1} = \sum_{j=1}^{m} \frac{A_j}{(1-\lambda_j L)} .$$

Then (2) and (4) imply

$$(5) \qquad \varepsilon_t = \sum_{j=1}^{m} A_j \sum_{k=0}^{\infty} \lambda_j{}^k x_{t-k}.$$

The geometric sums on the right side fail to converge if $|\lambda_j| > 1$. Imposing that $|\lambda_j| < 1$ for j = 1, . . ., m is necessary and sufficient for ε_t to lie in the space spanned by current and lagged x's. (This argument fails to reveal why when $|\lambda_j| = 1$ for some j, ε_t lies in the space spanned by current and lagged x_t's. When $|\lambda_j| = 1$ for some j, then although x_t possesses a moving average representation, it possesses no autoregressive representation. In this case, although ε_t is in the closure of the linear space spanned by $\{x_t, x_{t-1}, \ldots\}$, it cannot be expressed in the form $\sum_{j=0}^{\infty} w_j x_{t-j}$ for any sequence of w_j's, but only as the limit of a sequence of such expressions.)

To find a fundamental moving average representation, we note that (2) and (3) imply that $g_x(z) = \lambda_0{}^2(1-\lambda_1 z) \ldots (1-\lambda_m z)(1-\lambda_1 z^{-1}) \ldots (1-\lambda_m z^{-1})$. Then we equate this to $g_x(z)$ given by (1) to get

$$\sigma_u{}^2 a(z)a(z^{-1}) = \sigma_\varepsilon{}^2 d(z)d(z^{-1})$$

or

$$(6) \qquad \sigma_u{}^2 a(z)a(z^{-1}) = \sigma_\varepsilon{}^2 (1-\lambda_1 z) \ldots (1-\lambda_m z)(1-\lambda_1 z^{-1}) \ldots (1-\lambda_m z^{-1}).$$

Equation (6) asserts that $d(z)d(z^{-1})\sigma_\varepsilon^2$ is a symmetric <u>factorization</u> of $\sigma_u^2 a(z)a(z^{-1})$ in which the zeroes $\sigma_u^2 a(z)a(z^{-1})$ that are not inside the unit circle are placed in $d(z)$, while those that are not outside are put into $d(z^{-1})$. (Note that since $\sigma_u^2 a(z)a(z^{-1})$ evaluated at z_0 equals its value at z_0^{-1}, the zeroes of $\sigma_u^2 a(z)a(z^{-1})$ come in reciprocal pairs. Thus, it is reminiscent of the characteristic polynomial of an Euler equation in the undiscounted case).

The preceding tells us how to achieve a Wold moving average representation for a finite-order moving average process. First, find the zeroes of $g_x(z)$. By the reciprocal pairs properties of these roots, half will not be outside the unit circle, while half will not be inside the unit circle. (Excuse the cumbersome wording, which is designed to cover the case of roots on the unit circle.) Let $\lambda_1, \ldots, \lambda_m$ be the roots that are inside the unit circle. Then set

(7) $$d(L) = (1-\lambda_1 L) \ldots (1-\lambda_m L).$$

Then to find σ_u^2, solve equation (6) at $z = 1$ to get

(8) $$\sigma_\varepsilon^2 = \frac{g_x(1)}{d(1)^2}$$

Hansen and Sargent [, p. 102] give a quick but equivalent method of finding $d(L)$. Given $\sigma_u^2 d(z)d(z^{-1})$, the problem is to find $\sigma_\varepsilon^2 d(z)d(z^{-1})$, where the zeroes of $d(z)$ are not inside the unit circle. Let $z_1, \ldots, z_k$ be the zeroes of $a(z)$ that are outside the unit circle, where $0 \leq k \leq m$. Then $d(z)$ satisfies

(9) $$\frac{\sigma_\varepsilon}{\sigma_u} d(z) = a(z) \prod_{j=1}^{k} \frac{(1-z_j z)}{(z-z_j)} .$$

For example, suppose $x_t = (1+2L)u_t$, where $\sigma_u^2 = 1$. Then application of (9) gives the Wold moving average representation $x_t = (1+(1/2)L)u_t$, with $\sigma_\varepsilon = 2$.

Finding a Wold representation: An m^{th} order moving average, n^{th} order autoregression

The following problem is a useful input into solving an interesting class of "signal extraction" problems.

Consider a covariance stationary process x_t with representation

$$x_t = \frac{a(L)}{b(L)} u_t$$

where u_t is a (not necessary a fundamental) white noise and

$$b(L) = (1-\mu_1) \; . \; . \; . \; (1-\mu_n L), \; |\mu_j| < 1$$

$$a(L) = (1-\alpha_1 L) \; . \; . \; . \; (1-\alpha_m L)$$

Note that we assume that the zeroes of $b(z)$ are outside the unit circle, but that those of $a(z)$ are unrestricted. Our problem is to find a Wold moving average representation for x_t.

The solution of this problem is simply

$$x_t = \frac{d(L)}{b(L)} \varepsilon_t$$

where $d(L) = (1-\lambda_1 L) \; . \; . \; . \; (1-\lambda_m L)$, where $\lambda_1, \; . \; . \; ., \lambda_m$ are the zeroes of $a(z)a(z^{-1})$ that do not lie outside the unit circle, and where ε_t is the fundamental white noise for x_t, with variance σ_ε^2 given by

$$\sigma_\varepsilon^2 = \frac{a(1)^2 \sigma_u^2}{d(1)^2} .$$

In other words, the denominator polynomial b(L) is left unaltered while the methods of the preceding section are applied to factor the numerator polynomial. The reader should convince himself that this method delivers an ε_t process that is a white noise, and that lies in the linear space spanned by $\{x_t, x_{t-1}, \ldots\}$. This can be done by constructing an argument along the lines of the one in the preceding section, by assuming $|\lambda_j| < 1$ for $j = 1, \ldots, m$ and by premultiplying (10) by $b(L)/d(L)$.

Signal Extraction Problems

Let y_t be a covariance stationary stochastic process with m^{th} order moving average representation

$$y_t = a(L)u_t,$$

where $a(L) = \sum_{j=0}^{m} a_j L^j$, where

u_t is a white noise with variance σ_u^2 that is not necessarily fundamental for y_t. Suppose that x_t is the sum of y_t and an orthogonal serially uncorrelated white noise η_t with variance σ_η^2,

$$x_t = y_t + \eta_t$$

where $E\eta_t u_{t-s} = 0$ for all s.

Suppose that an agent observes $\{x_t, x_{t-1}, \ldots\}$ at t, and wishes to construct linear least squares forecasts of x's on the basis of this information set. To construct the linear least squares forecast for x_{t+k} given $\{x_t, x_{t-1}, \ldots\}$, one uses the Wiener-

Kolmogorov formula (52), which requires that a Wold moving average representation $x_t = d(L)\varepsilon_t$ be obtained for x_t.

To obtain the Wold representation for x_t, we simply use the method of section ____. In particular, the covariance generating function of x_t is

$$g_x(z) = a(z)a(z^{-1})\,\sigma_u^{\,2} + \sigma_\eta^{\,2}$$

We find the zeroes of $g_x(z)$, which come in reciprocal pairs, and prepare the factorization

$$g_x(z) = d(z)\;d(z^{-1})\;\sigma_\varepsilon^{\,2}$$

where the zeroes of $d(z) = (1-\lambda_1 z)\;.\;.\;.(1-\lambda_m z)$ do not lie inside the unit circle, and $\sigma_\varepsilon^{\,2}$ solves

$$\sigma_\varepsilon^{\,2} = \frac{g_x(1)}{d(1)^2}$$

The Wiener-Kolmogorov formula (52) can then be used, to calculate $P[x_{t+k} | x_t, x_{t-1}. . .]$.

Moving into a richer class of examples, we now let y_t be a process with mixed moving average, autoregressive representation

$$y_t = \frac{a(L)}{b(L)}\,u_t$$

where u_t is a white noise with variance $\sigma_u^{\,2}$, and

$$a(L) = (1-\alpha_1 L)\;.\;.\;.(1-\alpha_m L)$$

$$b(L) = (1-\mu_1 L)\;.\;.\;.(1-\mu_m L),\;\; |\mu_j| < 1$$

where the α_j's can be on either side of the unit circle. Suppose that x_t is the sum of y_t and a serially uncorrelated white noise η_t with variance σ_η^2,

$$x_t = y_t + \eta_t$$

where $E\eta_t \; u_{t-s} = 0$ for all s. Again we desire to find $P[x_{t+k}|x_t, x_{t-1}, \ldots]$, so we need to find a Wold representation for x_t. We use the method of section ___.

The covariance generating function of x is

$$g_x(z) = \frac{a(z)a(z^{-1})}{b(z)b(z^{-1})} \sigma_u^2 + \sigma_\eta^2 .$$

Taking the right hand side to a common denominator gives

$$g_x(z) = \frac{\sigma_u^2 a(z)a(z^{-1}) + \sigma_\eta^2 b(z)b(z^{-1})}{b(z)\; b(z^{-1})} \tag{11}$$

The numerator polynomial is of order p = max (n,m), and can be factored to be of the form

$$\sigma_u^2 a(z)a(z^{-1}) + \sigma_\eta^2 b(z)b(z^{-1}) = \sigma_\varepsilon^2 d(z)d(z^{-1}) \tag{12}$$

where

$$d(z) = (1-\lambda_1 L) \ldots (1-\lambda_p L), \; |\lambda_j| < 1 \; | \; j = 1 \ldots, p$$

and where σ_ε^2 solves

$$\sigma_\varepsilon^2 = \frac{\sigma_u^2 a(1)^2 + \sigma_\eta^2 b(1)^2}{d(1)^2} .$$

The Wold moving average representation for x_t is then

(13) $$x_t = \frac{d(L)}{b(L)} \varepsilon_t \ .$$

The Wiener-Kolmogorov formula can be applied to (13).

A famous application of the preceding analysis is due to Muth [ ]. Muth assumed that income x_t is the sum of a first order Markov process $[1/(1-\rho L)]u_t$, $|\rho| < 1$, and an uncorrelated white noise η_t. The agent's problem was to predict his future income. Setting $a(L) = 1$, $b(L) = (1-\rho L)$, we find that equation (12) becomes

$$\sigma_u^2 + \sigma_\eta^2(1-\rho z)(1-\rho z^{-1}) = \sigma_\varepsilon^2(1-\lambda_1 z)(1-\lambda_1 z^{-1}).$$

The expression on the left can be written

$$\rho z^{-1}\sigma_\eta^2[-z^2 + (\frac{\sigma_u}{\sigma_\eta^2\rho} + (\frac{1}{\rho} + \rho)) -1].$$

Applying the quadratic formula, and setting λ_1 equal to the root that is smaller in absolute value, we have

(14) $$\lambda_1 = \frac{1}{2} [(\frac{\sigma_u^2}{\sigma_\eta^2\rho}) +(\frac{1}{\rho} + \rho)] - \sqrt{[\frac{\sigma_u^2}{\sigma_\eta^2\rho} + (\frac{1}{\rho} + \rho)]^2 - 4}$$

The limiting value of λ_1 as ρ approaches 1 from below is

(15) $$\lambda_1 = 1 + \frac{1}{2} (\frac{\sigma_u^2}{\sigma_\eta^2}) - \sqrt{\frac{\sigma_u^2}{\sigma_\eta^2} (1 + \frac{1}{4} \frac{\sigma_u^2}{\sigma_\eta^2})} \quad ,$$

which is the expression obtained by Muth [ ]. Thus, we have that x_t has the first-order moving average, first order autoregressive representation

$$x_t = \left(\frac{1-\lambda_1 L}{1-\rho L}\right)\varepsilon_t,$$

where ε_t is a fundamental white noise for x_t with variance σ_ε^2 that solves

$$\sigma_\varepsilon^{\ 2} = \frac{\sigma_\varepsilon^{\ 2} + \sigma_\eta^{\ 2}(1-\rho)^2}{(1-\lambda_1)^2} .$$

The result from page ____ now applies with $\beta \equiv \rho$ and $\lambda_1 \equiv - a$. Thus we have

$$P_t \ x_{t+k} = [\rho^k(\rho-\lambda_1) \ / \ (1-\lambda_1 L)]x_t,$$

so that projections of future x's are a geometric average of past x's.

5. Signal Extraction With Dynamics

We now use the recursive projection formula (15') to solve a signal extraction problem that John Muth [ ] used to provide a rationalization for Milton Friedman's formula for permanent income. This will lead us to a version of the Kalman filter.

We consider the structure

$$(1) \qquad \theta_{t+1} = \rho\theta_t + \varepsilon_{t+1}$$

$$(2) \qquad z_t = c\theta_t + u_t$$

where ρ and c are scalars, ε_{t+1} is a random variable satisfying $E\,\varepsilon_t = 0$ for all t, $E\,\varepsilon_t^2 = \sigma_\varepsilon^2$ for all t, and $E\,\varepsilon_t\varepsilon_{t-s} = 0$ for $s \neq 0$. We assume that u_t is a random variable satisfying $E\,u_t = 0$ for all t, $E\,u_t^2 = \sigma_u^2$ for all t, $E\,u_t u_{t-s} = 0$ for $s \neq 0$ and $E\,u_t\varepsilon_s = 0$ for all t, s. Equation (1) states that θ_t is governed by a first-order linear stochastic difference equation, while equation (2) states that z_t is a linear combination of θ_t and a "noise" u_t. Given this structure, we imagine the following problem that is to be solved by an agent who knows the values of $(c,\rho,\sigma_u^2, \sigma_\varepsilon^2)$. At time t, the agent is imagined to see $(z_t,z_{t-1},\ldots,z_0)$, but not to have seen θ for any t. Thus θ_t is a "hidden variable." At time 0, before observing z_0, the agent is imagined to have an initial idea about the location of θ_0, which can be summarized by saying that he thinks it is distributed with mean $\hat{\theta}_0$ and variance about θ_0 of Σ_0. The agent's problem is to calculate $P[\theta_{t+1}|z_t,z_{t-1},\ldots,z_0]$. Using (15'), we shall derive a convenient recursive formula for this projection.

We define $\hat{\theta}_{t+1} = P[\theta_{t+1}|z_t, z_{t-1}, \ldots, z_0]$. In (15'), at $t \geq 1$, we let $y = \theta_{t+1}$, $\Omega = (z_{t-1}, z_{t-2}, \ldots, z_0)$, $x = z_t$. Then in light of (1)-(2), and using $P\varepsilon_{t+1}|z_t, z_{t-1}, \ldots z_0 = 0$ by virtue of (1)-(2) and the orthogonality conditions assumed for (ε_{t+1}, u_t), (15') becomes

$$P[\theta_{t+1}|z_t, z_{t-1}, \ldots, z_0] = P[\rho\theta_t|z_{t-1}, z_{t-2}, \ldots, z_0]$$

$$+ P[\rho\theta_t - P[\rho\theta_t|z_{t-1}, \ldots, z_0])|(z_t - P[z_t|z_{t-1}, \ldots z_0]).$$

or

$$(3) \qquad \hat{\theta}_{t+1} = \rho\hat{\theta}_t + P[\rho(\theta_t - \hat{\theta})|(c(\theta_t - \hat{\theta}_t) + u_t)].$$

where in the last line we use $z_t - P[z_t|z_{t-1}, \ldots z_0] = c\theta_t + u_t - c\hat{\theta}_t$. Let us define

$$(4) \qquad \Sigma_t = E(\theta_t - \hat{\theta}_t)^2.$$

Then we have that

$$(5) \qquad P[\rho(\theta_t - \hat{\theta}_t)|c(\theta_t - \hat{\theta}_t) + u_t] = K_t[c(\theta_t - \hat{\theta}_t) + u_t]$$

where

$$(6) \qquad K_t = \frac{c\ \rho\ \Sigma_t}{c^2\Sigma_t + \sigma_u^2}.$$

In deriving (6), we use the orthogonality condition $Eu_t(\theta_t - \hat{\theta}_t) = 0$ which follows from (1)-(2) and the orthogonality conditions imposed on (ε_{t+1}, u_t). Substituting (5) into (3) and rearranging gives

$$(7) \qquad \hat{\theta}_{t+1} = (\rho - K_t c)\hat{\theta}_t + K_t z_t.$$

Subtracting (7) from (1) and using (2) gives

$$\theta_{t+1} - \hat{\theta}_{t+1} = (\rho - K_t c)(\theta_t - \hat{\theta}_t) + \varepsilon_{t+1} - K_t u_t$$

Computing variances gives

$$\Sigma_{t+1} = (\rho - K_t c)^2 \Sigma_t + \sigma_\varepsilon^{\ 2} + K_t^{\ 2} \sigma_u^{\ 2}. \tag{8}$$

Equations (6), (7) and (8) are to be solved starting from the initial condition Σ_0 given. These three equations give a convenient recursive solution to our problem. The equations are a scalar verson of the famous "Kalman filter."

By analyzing the pair of difference equations (6), (8), it is possible to establish the following two properties of the solution. First, for any value of ρ and for any value $c \neq 0$, starting the system from any $\Sigma_0 \geq 0$ leads to a Σ_t sequence that converges as $t \to \infty$. Second, for the same range of values of ρ and c, the parameter $(\rho - Kc)$, where $K = \lim_{t \to \infty} K_t$, is less than unity in absolute value. This implies that for the infinite history filtering problem, in which the agent is imagined to form $\hat{\theta}_{t+1} = P[\theta_{t+1} | z_t, z_{t-1}, \ldots]$ by projecting on an infinite record of current and past z's, the solution can be represented by the time invariant equation

$$\hat{\theta}_{t+1} = (\rho - Kc)\hat{\theta}_t + Kz_t. \tag{9}$$

where $|(\rho - Kc)| < 1$, where K is the unique stationary solution of (6) and (8) that is associated with a stationary solution for which $\lim \Sigma_t = \Sigma > 0$. Equation (9) can be solved to give a version of Friedman and Cagan's formula

(10) $$\hat{\theta}_{t+1} = K \sum_{j=0}^{\infty} (\rho - Kc)^j z_{t-j}.$$

Muth considered the case in which $c = 1$, $\rho = 1$. In this case, (10) becomes exactly the adaptive expectations mechanism that was used by Friedman and Cagan.

Note that the orthogonality conditions imposed on the (ε, u) process imply that

(11) $$P[\theta_{t+j} | z_t, z_{t-1}, \ldots] = \rho^{j-1} [\rho - Kc] \hat{\theta}_{t+1} \text{ for } j \geq 1.$$

Additional Exercises

1. Suppose in (1)-(2), that $c \neq 0$. Prove that starting from $\Sigma_0 = 0$, iterations on (8) produce a convergent sequence. (Hint: first prove that starting from $\Sigma_0 = 0$, iterations on (8) lead to a monotone sequence. Then argue that this sequence is bounded by producing a naive estimator compared to which the linear projection must give a lower variance of estimate.)
2. Using the results of exercise (1), prove that starting from $\Sigma_0 = 0$, the stationary value of the parameter $(\rho - Kc)$ must be less than unity in absolute value. (Hint: you have already proved that Σ_t converges starting from Σ_0.)
3. Prove that the stationary value of Σ is independent of Σ_0, and that (6)-(8) have a unique stationary solution with $\Sigma \geq 0$.
4. Substitute stationary values into (6) and (8), and eliminate K to get a quadratic equation in Σ. Argue that this equation resembles an Euler equation for an undiscounted quadratic optimization problem.

Evaluating the Inverse z-Transform

In the previous section, we saw in the corollary to the Riesz-Fischer theorem that given a square summable sequence $\{c_j\}$, there exists a function $g(z)$, defined as

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

which is well defined at least on the unit circle ($z=e^{iw}$, $w\epsilon [0,2\pi]$). The function $g(z)$ is often called the "z-transform" of the sequence c_j. The function g maps points z in the complex plane into points $g(z)$ in the complex plane. Furthermore, we saw that the c_j can be recovered from the function $g(z)$ by the inversion formula

$$\text{(a)} \qquad c_k = \frac{1}{2\pi i} \int_\Gamma g(z) z^{-k-1} dz, \; k = 0, \pm 1, \pm 2, \ldots$$

where the integral is a contour integral in the complex plane, and Γ denotes the unit circle. In this section, we give a pair of simple formulas for evaluating the integral on the right side of (a). Virtually no knowledge of complex analysis is required to use the formulas.

As a prelude to giving this formula, we need to explain two concepts: that of a pole of the complex valued function $g(z)$, and the residue that is associated with each pole.

Roughly speaking, a pole is a point in the complex plane, say z_0, such that $g(z)$ approaches infinity as z approaches z_0. In this book, we work almost entirely with functions $g(z)$ that are rational, that is, ratios of finite order polynomials in z. For this reason, the following test for poles of order m sufficient to identify all of the poles of a function $g(z)$:

Test for poles: Given a function $g(z)$, suppose that for some positive integer m the function

(b) $$\phi(z) = (z-z_0)^m g(z)$$

can be defined[1/] so that $\phi(z_0) \neq 0$. Then $g(z)$ has a pole of order m at z_0.

To illustrate the concept of a pole, consider the important case in which $g(z)$ is a rational function, $g(z) = a(z)/b(z)$ where $a(z)$ and $b(z)$ are finite order polynomials in z with no zeroes in common. Then according to the above test for poles, the poles of $g(z)$ are simply the zeroes of $b(z)$. For one example, letting $g(z) = 1/(1-\lambda z)^r$, r an integer, we find that $g(z)$ has a pole of order r at $z = \lambda^{-1}$.

We now turn to define the residue associated with a pole at z_0.

Definition of residue: Suppose that $g(z)$ has a pole of order m at $z = z_0$. Define the function $\phi(z) = (z-z_0)^m g(z)$. Then the residue associated with z_0 is defined as

(c) $$\text{res}(z_0) = \frac{\phi^{(m-1)}(z_0)}{(m-1)!}$$

where $\phi^{(m-1)}(z_0)$ is the $(m-1)^{th}$ order derivative of ϕ evaluated at z_0. In the important special case in which $m = 1$, formula (c) reduces to $\text{res}(z_0) = \phi(z_0)$ or

(d) $$\text{res}(z_0) = \lim_{z \to z_0} (z-z_0)g(z).$$

We now give two convenient formulas for evaluating the inversion integral

$$c_j = \frac{1}{2\pi i} \int_\Gamma g(z) z^{-j-1} dz.$$

The formulas are given by $c_j = \frac{1}{2\pi i} \int_\Gamma g(z) z^{-j-1} dz$

(e)
$$= \left[\text{sum of residues of } g(z^{-1})z^{j-1} \text{ at poles inside unit circle}\right]$$
$$= \left[\text{sum of residues of } g(z)z^{-j-1} \text{ at poles inside unit circle}\right]$$

It is a good idea to use whichever of the two branches of (e) is most convenient to avoid the appearance of poles of order greater than one at zero.2/

We now illustrate the utility of formulas (e) with some examples. For our first example, we take $g(z) = 1/(1-\lambda z)$ with $|\lambda| < 1$. Using (e), we have that for $j > 0$, c_j = sum of residues of $z^{j-1}/(1-\lambda z^{-1})$ inside the unit circle. For $j > 0$, the function $z^{j-1}/(1-\lambda z^{-1})$ has a single pole of order one at $z = \lambda$, with residue given by $\lim_{z \to \lambda} (z-\lambda) z^{j-1}/(1-\lambda z^{-1}) = \lambda^j$.
Therefore, for $j > 0$, $c_j = \lambda^j$. For $j < 0$ we use the second branch of (e) and find that the function $z^{-j-1}/(1-\lambda z)$ has no poles inside the unit circle. Therefore, $c_j = 0$ for $j < 0$. Finally, for $j = 0$, we use the second branch of (e), and find that the function $z^{-j-1}/(1-\lambda z)$ has a pole of order one at $z = 0$, which is the only pole inside the unit circle. The residue associated with this pole is $\lim_{z \to 0} (z-0) z^{-1}(1-\lambda z)^{-1} = 1$. Therefore $c_0 = 1$. Thus we have

$$c_j = \begin{cases} \lambda^j & j \geq 0 \\ 0 & j < 0. \end{cases}$$

Of course, these results could be more easily obtained simply by expanding $1/(1-\lambda z)$ in a geometric sum. However, for some of the more complicated examples to be described below, the residue calculations are quicker, then such an alternative method.

As a second example, consider the covariance generating function of the n^{th} order autoregressive process

$$g_y(z) = \frac{1}{(1-\lambda_1 z)\ldots(1-\lambda_n z)(1-\lambda_1 z^{-1})\ldots(1-\lambda_n z^{-1})}$$

$$= \frac{z^n}{(1-\lambda_1 z)\ldots(1-\lambda_n z)(z-\lambda_1)\ldots(z-\lambda_n)}$$

where $|\lambda_j| < 1$ for $j = 1,\ldots,n$.

Using formula (e) to evaluate

$$c_y(\tau) = \frac{1}{2\pi i}\int_\Gamma g_y(z) z^{-\tau-1} dz, \text{ we have}$$

$$c_y(\tau) = \text{sum of residues of } \frac{z^{n-\tau-1}}{(1-\lambda_1 z)\ldots(1-\lambda_n z)(z-\lambda_1)\ldots(z-\lambda_n)}$$

(f) at poles inside the unit circle

$$= \text{sum of residues of } \frac{z^{n+\tau-1}}{(1-\lambda_1 z)\ldots(1-\lambda_n z)(z-\lambda_1)\ldots(z-\lambda_n)}$$

at poles inside the unit circle.

It is convenient to use the first line of (f) for $\tau \leq 0$ and the second line for $\tau \geq 0$, in order to avoid poles of multiple order at $z = 0$. In each case, the function has poles of order one at $\lambda_1,\ldots, \lambda_n$. For $\tau \geq 0$ residue at pole λ_j is readily found to be

$$\text{res}\ (\lambda_j) = \frac{\lambda_j^{n+\tau-1}}{\prod_{k=1}^{n} (1-\lambda_k\lambda_j) \prod_{\substack{k=1\\k\neq j}}^{n} (\lambda_j-\lambda_k)}, \quad \tau \geq 0.$$

It follows that

$$\text{(g)} \qquad c_y(\tau) = \sum_{j=1}^{n} \frac{\lambda_j^{n+|\tau|-1}}{\prod_{k=1}^{n} (1-\lambda_k\lambda_j) \prod_{\substack{k=1\\k\neq j}}^{n} (\lambda_j-\lambda_k)}$$

where the absolute value sign in (g) follows either the symmetry of the covariogram or from pursuing the implication of the second line of (f).

Additional Exercises

(i). Let y_t be a mixed moving average, autoregressive process $y_t = (B(L)/A(L))\varepsilon_t$, where ε_t is a white noise with unit variance, $B(L) = \prod_{j=1}^{m} (1-\mu_j L)$, and $A(L) = \prod_{j=1}^{n} (1-\lambda_j L)$, $|\lambda_j| < 1$ for $j = 1,\ldots, n$, and where $\lambda_j \neq \lambda_i$ for $\substack{j=1 \\ i\neq j}$, $\lambda_j \neq \mu_k$ for all $j = 1,\ldots,n$, $k = 1,\ldots, m$, and $m \leqslant n$. The autovariance generating function for y is $g_y(z) = B(z)B(z^{-1})/A(z)A(z^{-1})$. Use formula (e) to establish the formula

$$c_y(\tau) = \sum_{s=1}^{n} \frac{\lambda_s^{n+|\tau|-m-1} \prod_{j=1}^{m} (1-\mu_j \lambda_s)(\lambda_s-\mu_j)}{\prod_{j=1}^{n} (1-\lambda_j\lambda_s) \prod_{\substack{j=1 \\ j\neq s}}^{n} (\lambda_s-\lambda_j)} .$$

(ii). Let b(L) be the polynomial in the lag operator $b(L) = (1+\mu L)/(1-\lambda L) = \sum_{j=-\infty}^{\infty} b_j L^j$ where $|\lambda| < 1$. Use formula (e) to establish that

$$b_j = \begin{cases} 0 & j < 0 \\ 1 & j = 0 \\ \lambda^j + \mu\lambda^{j-1} & j \geqslant 1. \end{cases}$$

(iii). Consider the generating function of the second-order Solow-Pascal lag distribution $w(z) = 1/(1-\lambda z)^2$, $|\lambda| < 1$. Use formulas (c) and (e) to evaluate the coefficients of the lag distribution. Compare your results with equation of chapter IX.

Footnotes

[1] The function $\phi(z)$ must also be "analytic" at z_0, which means that its derivative exists at z_0 and at each point in same neighborhood of z_0. In our examples, this requirement will routinely be satisfied.

[2] The second representation in (e) is a standard in texts in complex analysis (e.g., see Churchill [ ]). The first is derived simply from the second as follows. Notice that $g(z^{-1}) = \sum_{n=-\infty}^{\infty} c_n z^{-n} = \sum_{j=-\infty}^{\infty} c_{-j} z^{j} = \sum_{j=-\infty}^{\infty} d_j z^j$ with $d_j = c_{-j}$. Using the second representation in (e) to calculate d_j, it follows that c_j = sum of residues of $g(z^{-1})z^{j-1}$ at poles inside the unit circle.

Predicting Geometric Distributed Leads

It is important to know the solution of the following problem in order to use a variety of linear rational expectations models. Let x_t be a covariance stationary stochastic process with Wold moving average representation

(1) $$x_t = c(L)\varepsilon_t, \quad c_0 = 1$$

where ε_t is a fundamental white noise for x and $c(L) = \sum_{j=0}^{\infty} c_j L^j$ is square summable. We further assume that c(L) has an inverse a(L) $= c(L)^{-1}$ which is one-sided in nonnegative powers of L and square summable. Thus, x_t has the autoregressive representation

(2) $$a(L)x_t = \varepsilon_t$$

where $a(L) = 1 - a_0 L - a_2 L^2 - \ldots.$

We want to calculate the following linear projection

(3) $$y_t = P[\sum_{j=0}^{\infty} \lambda^j x_{t+j} | x_t, x_{t-1}, \ldots] \equiv P_t \sum_{j=0}^{\infty} \lambda^j x_{t+j}.$$

where $|\lambda| < 1.$

Projections of such geometric distributed leads occur in a variety of linear rational expectations model. We begin by noting that y_t defined by (3) satisfies the stochastic difference equation

(4) $$y_t = \lambda P_t y_{t+1} + x_t.$$

That is, y_t is the stationary solution of the difference equation (3) as can be verified by repeated substitution in (4). We seek expressions for y_t of the forms

(5) $$y_t = g(L)x_t$$

and

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

where $d(L) = \sum_{j=0}^{\infty} \lambda_j L^j$, $g(L) = \sum_{j=0}^{\infty} g_j L^j$,

$\sum_{j=0}^{\infty} g_j^2 < + \infty$, $\Sigma d_j^2 < + \infty$.

We know that representation (5) exists by definition, and therefore that $d(L) = g(L)c(L)$ also exists. That is, a representation of the form (6) exists because $\{x_t, x_{t-1}, \ldots\}$ and $\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$ span the same space.

We shall solve for $d(L)$ using (4) and prediction theory. Using (6), we have that

$$P_t y_{t+1} = [\frac{d(L)}{L}]_+ \varepsilon_t \text{ or}$$

$$P_t y_{t+1} = [\frac{d(L)}{L} - \frac{d_0}{L}]\varepsilon_t.$$

Substituting this and (4) and (5) into (4) gives

$$d(L)\ \varepsilon_t = \lambda[\frac{d(L)}{L} - \frac{d_o}{L}]\varepsilon_t + c(L)\varepsilon_t.$$

Since this equation holds for all ε_t realizations, it implies, after rearranging, that

$$(1-\lambda L^{-1})d(L) = c(L) - \lambda d_0 L^{-1},$$

an equation that we desire to solve for $d(L)$ as a function of $c(L)$. We determine d_0 by evaluating the above equation at $L = \lambda$, to get $c(\lambda) = d_0$. Using this value for λd_0 gives,

(7) $$d(L) = \frac{c(L) - \lambda c(\lambda)L^{-1}}{1-\lambda L^{-1}}.$$

Using $g(L) = d(L)c(L)^{-1}$ and $c(L) = a(L)$, we get

$$(8) \qquad g(L) = \frac{1 - \lambda a(\lambda)^{-1} a(L) L^{-1}}{1-\lambda L^{-1}}.$$

For the case in which $a(L)$ is an r^{th} order polynomial $a(L) = 1 - \sum_{j=1}^{r} a_j L^j$, Hansen and Sargent [ ] show by using polynomial long division that (8) can be expressed

$$(9) \qquad g(L) = a(\lambda)^{-1}[1+\sum_{j=1}^{r-1} (\sum_{k=j+1}^{r} \lambda^{k-j} a_t) L^j]$$

so that $g(L) = \sum_{j=0}^{r-1} g_j L^j$, with $g_o = a(\lambda)^{-1}$, $g_0 = a(\lambda)^{-1} \sum_{k=j+1}^{r-1} \lambda^{k-j} \alpha_k$

for $j = 1,\ldots,r - 1$.

Various versions of formulas (7), (8), and (9) were originally derived in papers by Saracoglu and Sargent [ ], Hansen and Sargent [ ], and Futia [ ].

Exercise:

_____. Assume that m_t is covariance stationary and has an autoregressive representation $a(L)m_t = \varepsilon_t$ where ε_t is a fundamental white noise for m_t, and $a(L) = 1 - a_1L - \dots - a_rL^r$.

a. Define a state vector x_t and a unit vector e, and use it to express the law of motion for $m_t = ex_t$ in the first-order vector form $x_t = Ax_{t-1} + e_t$ where e_t is a vector white noise.

b. Use the formula (   ) to derive the formula (H-S) by inverting $(I-\lambda A)$, taking into account the many zeroes in $(I-\lambda A)$.

Insert F

This difference equation can be rewritten as

$$p_t = \left(\frac{-\alpha}{1-\alpha}\right) P_t p_{t+1} + \left(\frac{1}{1-\alpha}\right) m_t$$

or

$$p_t = \lambda P_t p_{t+1} + (1-\lambda) m_t$$

where $\lambda = -\alpha/1-\alpha$, which implies that $o < \lambda < 1$ since $\alpha < 0$. The stationary solution of the above difference equation obeys 27/

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

Let us assume that m_t has the autoregressive representation

$$a(L) m_t = \varepsilon_t$$

where ε_t is fundamental for m, and $a(L) = 1 - a_1 L \ldots a_r L^r$. Then from formula (   ) of the preceding section we have that (1) implies

(i) $$p_t = (1-\lambda) a(\lambda)^{-1} [1 + \sum_{j=1}^{r-1} (\sum_{k=j+1}^{r} \lambda^{k-j} a_k) L^j] m_t$$

(ii) $$a(L) m_t = \varepsilon_t.$$

These two equations express how the stochastic process for p_t is a function of the stochastic process of m_t. Notice that p_t depends on m_t, $m_{t-1}, \ldots, m_{t-r+1}$ via coefficients that partly reflect the stochastic process (ii) that governs m_t. As an example, we set $a(L) = 1 - a_1 L - a_2 L^2 - a_3 L^3$. Then (i) and (ii) become

$$p_t = (1-\lambda)(1-a_1\lambda-a_2\lambda^2-a_3\lambda^3)^{-1} [1+(a_2\lambda+a_3\lambda^2)L + (a_3\lambda)L^2] m_t$$

Insert G

We have assumed that x_t has the autoregressive representation $a(L)x_t = \varepsilon_t$. Now by using methods similar to those used to derive (   ), it can be established that

$$P_{t-1} \sum_{j=0}^{\infty} \lambda^j x_{t+j} = \left(\frac{L^{-1}I - L^{-1}a(\lambda)^{-1}a(L)}{1-\lambda L^{-1}}\right) x_{t-1}.$$

Substituting this and (   ) into (60) we have the following formula for the equilibrium stochastic process for price p_t as a function of the x_t process:

$$p_t = \lambda p_{t-1} + \frac{1}{1 + \beta - \alpha\lambda} \left\{\alpha^{-1}\lambda(\gamma+\alpha)\left[\frac{L^{-1}I - L^{-1}a(\lambda)^{-1}a(L)}{1-\lambda L^{-1}}\right] x_{t-1} - \left(\frac{1-\lambda a(\lambda)^{-1}a(L)L^{-1}}{1-\lambda L^{-1}}\right) x_t\right\}$$

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

Insert H

for the log of the price level of the form

$$(\quad) \qquad p_t = (1-\lambda) \sum_{j=0}^{\infty} \lambda^j P_t m_{t+j}$$

where $\lambda = -\alpha/(1-\alpha)$, and where m_t is the log of the money supply. Suppose that m_t is the first element of a vector x_t that evolves according to $x_t = Ax_{t-1} + \varepsilon_t$ where ε_t is a vector white noise. Let e be the unit vector that validates our writing $m_t = ex_t$. Then substituting (69) into the above solution for p_t gives

$$p_t = (1-\lambda)e \left(\sum_{j=0}^{\infty} \lambda^j A^j \right) x_t.$$

If the eigenvalues of A are bounded by $1/\lambda$ is modulus,[28] then we have that $\sum_{j=0}^{\infty} \lambda^j A^j = (I-\lambda A)^{-1}$. Therefore, our solution can be represented

$$p_t = (1-\lambda)e(I-\lambda A)^{-1} x_t$$

$$x_t = Ax_{t-1} + \varepsilon_t.$$

Two comments about this derivation are in order. First, in the special case in which only lagged m's appear in the x_{t-1} vector, the above formula is equivalent with formula ( ) on page ( ) [H-S formula with g(L)]. In fact that formula could be derived from the above one simply by explicitly inverting $(I-\lambda A)$.

Second, we notice from ( ) that not only lagged m's but also any other variables that appear in the vector x_t also enter the equation ( ) for p_t. Thus, any variables that help predict future m's end up in the equation (   ) expressing p_t as a function of current and lagged variables. We shall expand upon this shortly.

$$m_t = a_1 m_{t-1} + a_2 m_{t-2} + a_3 m_{t-3} + \varepsilon_t.$$

1. Introduction

In Chapter IX, we studied linear difference equations of the form

$$(1) \qquad (1-a_1-\ldots-a_n L^n)y_t = x_t,$$

where $\{x_t\}_{t=-\infty}^{\infty}$ was taken to be a known sequence. We studied how to find the class of sequences $\{y_t\}$ that satisfy the difference equation and a set of prescribed boundary conditions on the $\{y_t\}$ sequence. Such a $\{y_t\}$ sequence was said to solve the difference equation.

The present chapter studies linear difference equations of the form (1) in which, rather than being a sequence of known numbers, $\{x_t\}$ is a sequence of independently and identically distributed random variables with known variance and mean. With this choice of mechanism for generating $\{x_t\}$, equation (1) is called a linear stochastic difference equation. A solution of such a difference equation is a sequence of random variables $\{y_t\}$. A sequence of random variables is called a stochastic process. While the x_t sequence is by assumption a stochastic process consisting of random variables that are independently and identically distributed over time, the y_t process that solves (1) will in general be correlated over time. That is, while the $\{x_t\}$ process by assumption satisfies $E(x_t - Ex_t)(x_{t+s} - Ex_{t+s}) = 0$ for $s \neq 0$, for the y_t process in general $E(y_t - Ey_t)(y_{t+s} - Ey_{t+s}) \neq 0$ for $s \neq 0$. One way to characterize the solution of the difference equation is to summarize the second moments of the $\{y_t\}$ process and to describe how they depend on the a_j's of (1).

Stochastic difference equations provide a natural tool for interpreting and modeling economic time series. Macroeconomists spend much of their time interpreting sample first and second moments of observed time series. For example, for an observed sample on two variables $(y_t, z_t \; t = 1, \ldots, T)$ we often calculate various of the sample moments

$$T^{-1} \sum_{t=1}^{T} y_t, \; T^{-1} \sum_{t=1}^{T} z_t, \; (T-k)^{-1} \sum_{t=k+1}^{T} y_t \, y_{t-k},$$

$$(T-k)^{-1} \sum_{t=k+1}^{T} y_t \; x_{t-k} \quad \text{and} \quad (T-k)^{-1} \sum_{t=k+1}^{T} x_t \; y_{t-k}$$

for various values of k. It is convenient to adopt a mathematical context in which these sample moments can be regarded as estimators of the population moments Ey_t, Ez_t, $Ey_t \, y_{t-k}$, $Ey_t \, x_{t-k}$, and $Ex_t \, y_{t-k}$, respectively, estimators which converge to these population moments as $T \to \infty$. Linear stochastic difference equations provide such a mathematical context. In studying how to solve stochastic difference equations, one of our intermediate goals is to learn how the coefficients a_j of (1) can be chosen in order to make the implied pattern of population moments $Ey_t \, y_{t-k}$ resemble $(T-k)^{-1} \sum_{t=k+1}^{T} y_t \, y_{t-k}$ as the lag k is varied.

Stochastic processes provide a natural context in which to formulate the problem of prediction. At time t, suppose that observations on a stochastic process $(y_t, y_{t-1}, y_{t-2} \ldots)$ have not yet been revealed. Suppose that the moments Ey_t and $Ey_t \, y_{t-k}$ are known for all t and k. Then what is the best way to predict $(y_{t+1}, y_{t+2}, \ldots)$ as a linear function of $(y_t, y_{t-1}, \ldots)$? This linear prediction problem was solved by Wiener and Kolmogorov in the late 1930s.

The linear prediction problem is of interest to macroeconomists for at least two reasons. First, macroeconomists are interested in modeling the behavior of agents who are operating in dynamic and uncertain contexts. Typically, the hypothesis of utility or profit maximization ends up confronting those agents with some version of a prediction problem that they must solve in order best to achieve their objective. As we shall see, by using prediction theory, it is possible to extend the solutions of the quadratic dynamic optimization problems that were encountered in Chapter IX to the case in which the forcing functions are stochastic processes whose future values are not known at the time when decisions must be made. Thus, prediction theory is an important tool in determining optimizing behavior under uncertainty.

Second, macroeconomists are interested in using their own models of economic time series (often a collection of estimated a_j's in (1) or estimated moments $Ey_t \, y_{t-k}$) in order to predict the future conditional on the past. When the econometric model occurs in the form of a vector version of (1), it is said to be a <u>vector autoregression</u>. Linear prediction theory applies directly to such a model.

One of the goals of much recent work in rational expectations economics has been to create models whose equilibria are vector stochastic difference equations. In these models, the outcome of the interaction of a collection of purposeful agents is a stochastic process for, say, prices and quantities whose evolution can be described by a (vector) stochastic difference equation. We shall study versions of such models in which the equilibria are described by linear stochastic difference equations, i.e., vector versions of (1). In such models, some of the a_j's become interpretable in terms of the purposeful behavior of the agents in the model; that is, they are functions of the parameters of people's objective functions and constraints. One goal of this line of research is to acquire the ability to predict how the equilibrium stochastic process (or difference equation) would change in response to hypothetical changes in particular aspects of the environment confronting the agents in the model.

The idea that low order linear stochastic difference equations could provide a useful model for business cycles can be traced back at least as far as Slutsky (1937) and Frisch (1933). We have seen in Chapter IX that low order nonstochastic linear difference equations with no forcing functions present (i.e., x_t = 0 for all t in (1)) result in solutions for y_t that are "smooth," being the weighted sum of a small number of geometric sequences. Such smooth sequences do not resemble observed economic time series. However, if a sufficiently erratic forcing sequence $\{x_t\}$ occurs in (1), the resulting $\{y_t\}$ sequence that solves (1) can be sufficiently erratic that it resembles observed economic time series. The idea of Slutsky was to make the $\{x_t\}$ sequence sufficiently erratic by choosing it as the realization of a sequence of independently and identically distributed random variables. The resulting realizations of the $\{y_t\}$ sequence that solved*/ (1)

---

*/Here "solve" refers to the ordinary sense used in Chapter IX of finding a $\{y_t\}$ sequence that satisfies (1) given the realization of the {xt} sequence.

would be erratic enough to resemble observed time series. As we shall see, even first-order stochastic linear difference equations ($n = 1$ in (1)) can generate realizations that look like observed economic time series. Furthermore, the hypothesis that $\{x_t\}$ is a sequence of independently and identically distributed random variables in general implies that the future values (y_{t+1}, y_{t+2}, ...) are at best imperfectly predictable from past values (y_t, y_{t-1}, ...). It is desirable to have models in which both economic agents and econometricians confront uncertainty in this sense. This is one major reason that Slutsky's idea was adopted early on in dynamic econometrics, and why it has been retained and expanded upon in work on rational expectations.