Chapter 2.
Time series

2.1. Two workhorses

This chapter describes two tractable models of time series: Markov chains and first-order stochastic linear difference equations. These models are organizing devices that put particular restrictions on a sequence of random vectors. They are useful because they describe a time series with parsimony. In later chapters, we shall make two uses each of Markov chains and stochastic linear difference equations: (1) to represent the exogenous information flows impinging on an agent or an economy, and (2) to represent an optimum or equilibrium outcome of agents' decision making. The Markov chain and the first-order stochastic linear difference both use a sharp notion of a state vector. A state vector summarizes the information about the current position of a system that is relevant for determining its future. The Markov chain and the stochastic linear difference equation will be useful tools for studying dynamic optimization problems.

2.2. Markov chains

A stochastic process is a sequence of random vectors. For us, the sequence will be ordered by a time index, taken to be the integers in this book. So we study discrete time models. We study a discrete state stochastic process with the following property:

MARKOV PROPERTY: A stochastic process $\{x_t\}$ is said to have the *Markov property* if for all $k \geq 1$ and all t,

$$\text{Prob}\left(x_{t+1}|x_t, x_{t-1}, \ldots, x_{t-k}\right) = \text{Prob}\left(x_{t+1}|x_t\right).$$

We assume the Markov property and characterize the process by a *Markov chain.* A time-invariant Markov chain is defined by a triple of objects, namely, an n-dimensional state space consisting of vectors $e_i, i = 1, \ldots, n$, where e_i is an $n \times 1$

unit vector whose ith entry is 1 and all other entries are zero; an $n \times n$ *transition matrix* P, which records the probabilities of moving from one value of the state to another in one period; and an $(n \times 1)$ vector π_0 whose ith element is the probability of being in state i at time 0: $\pi_{0i} = \text{Prob}(x_0 = e_i)$. The elements of matrix P are

$$P_{ij} = \text{Prob}\left(x_{t+1} = e_j | x_t = e_i\right).$$

For these interpretations to be valid, the matrix P and the vector π must satisfy the following assumption:

ASSUMPTION M:

a. For $i = 1, \ldots, n$, the matrix P satisfies

$$\sum_{j=1}^{n} P_{ij} = 1. \tag{2.2.1}$$

b. The vector π_0 satisfies

$$\sum_{i=1}^{n} \pi_{0i} = 1.$$

A matrix P that satisfies property (2.2.1) is called a *stochastic matrix.* A stochastic matrix defines the probabilities of moving from each value of the state to any other in one period. The probability of moving from one value of the state to any other in *two* periods is determined by P^2 because

$$\begin{aligned}
&\text{Prob}\left(x_{t+2} = e_j | x_t = e_i\right) \\
&= \sum_{h=1}^{n} \text{Prob}\left(x_{t+2} = e_j | x_{t+1} = e_h\right) \text{Prob}\left(x_{t+1} = e_h | x_t = e_i\right) \\
&= \sum_{h=1}^{n} P_{ih} P_{hj} = P_{ij}^{(2)},
\end{aligned}$$

where $P_{ij}^{(2)}$ is the i,j element of P^2. Let $P_{i,j}^{(k)}$ denote the i,j element of P^k. By iterating on the preceding equation, we discover that

$$\text{Prob}\left(x_{t+k} = e_j | x_t = e_i\right) = P_{ij}^{(k)}.$$

The unconditional probability distributions of x_t are determined by

$$\begin{aligned}
\pi_1' &= \text{Prob}\left(x_1\right) = \pi_0' P \\
\pi_2' &= \text{Prob}\left(x_2\right) = \pi_0' P^2 \\
&\ \ \vdots \\
\pi_k' &= \text{Prob}\left(x_k\right) = \pi_0' P^k,
\end{aligned}$$

where $\pi_t' = \text{Prob}(x_t)$ is the $(1 \times n)$ vector whose ith element is $\text{Prob}(x_t = e_i)$.

2.2.1. Stationary distributions

Unconditional probability distributions evolve according to

$$\pi_{t+1}' = \pi_t' P. \tag{2.2.2}$$

An unconditional distribution is called *stationary* or *invariant* if it satisfies

$$\pi_{t+1} = \pi_t,$$

that is, if the unconditional distribution remains unaltered with the passage of time. From the law of motion (2.2.2) for unconditional distributions, a stationary distribution must satisfy

$$\pi' = \pi' P \tag{2.2.3}$$

or

$$\pi' (I - P) = 0.$$

Transposing both sides of this equation gives

$$(I - P')\, \pi = 0, \tag{2.2.4}$$

which determines π as an eigenvector (normalized to satisfy $\sum_{i=1}^n \pi_i = 1$) associated with a unit eigenvalue of P'.

The fact that P is a stochastic matrix (i.e., it has nonnegative elements and satisfies $\sum_j P_{ij} = 1$ for all i) guarantees that P has at least one unit eigenvalue, and that there is at least one eigenvector π that satisfies equation (2.2.4). This stationary distribution may not be unique because P can have a repeated unit eigenvalue.

Example 1. A Markov chain

$$P = \begin{bmatrix} 1 & 0 & 0 \\ .2 & .5 & .3 \\ 0 & 0 & 1 \end{bmatrix}$$

has two unit eigenvalues with associated stationary distributions $\pi' = [\,1 \quad 0 \quad 0\,]$ and $\pi' = [\,0 \quad 0 \quad 1\,]$. Here states 1 and 3 are both *absorbing* states. Furthermore, any initial distribution that puts zero probability on state 2 is a stationary distribution. See exercises 1.10 and 1.11.

Example 2. A Markov chain

$$P = \begin{bmatrix} .7 & .3 & 0 \\ 0 & .5 & .5 \\ 0 & .9 & .1 \end{bmatrix}$$

has one unit eigenvalue with associated stationary distribution $\pi' = [\,0 \quad .6429 \quad .3571\,]$. Here states 2 and 3 form an *absorbing subset* of the state space.

2.2.2. *Asymptotic stationarity*

We often ask the following question about a Markov process: for an arbitrary initial distribution π_0, do the unconditional distributions π_t approach a stationary distribution

$$\lim_{t\to\infty} \pi_t = \pi_\infty,$$

where π_∞ solves equation (2.2.4)? If the answer is yes, then does the limit distribution π_∞ depend on the initial distribution π_0? If the limit π_∞ is independent of the initial distribution π_0, we say that the process is *asymptotically stationary with a unique invariant distribution*. We call a solution π_∞ a *stationary distribution* or an *invariant distribution* of P.

We state these concepts formally in the following definition:

DEFINITION: Let π_∞ be a unique vector that satisfies $(I - P')\pi_\infty = 0$. If for all initial distributions π_0 it is true that $P^{t\prime}\pi_0$ converges to the same π_∞, we say that the Markov chain is asymptotically stationary with a unique invariant distribution.

The following theorems can be used to show that a Markov chain is asymptotically stationary.

THEOREM 1: Let P be a stochastic matrix with $P_{ij} > 0 \ \forall (i,j)$. Then P has a unique stationary distribution, and the process is asymptotically stationary.

THEOREM 2: Let P be a stochastic matrix for which $P_{ij}^n > 0 \ \forall (i,j)$ for some value of $n \geq 1$. Then P has a unique stationary distribution, and the process is asymptotically stationary.

The conditions of theorem 1 (and 2) state that from any state there is a positive probability of moving to any other state in 1 (or n) steps.

2.2.3. Expectations

Let $\overline{y}$ be an $n \times 1$ vector of real numbers and define $y_t = \overline{y}' x_t$, so that $y_t = \overline{y}_i$ if $x_t = e_i$. From the conditional and unconditional probability distributions that we have listed, it follows that the unconditional expectations of y_t for $t \geq 0$ are determined by $Ey_t = (\pi_0' P^t)\overline{y}$. Conditional expectations are determined by

$$E\left(y_{t+1}|x_t = e_i\right) = \sum_j P_{ij}\overline{y}_j = \left(P\overline{y}\right)_i \tag{2.2.5}$$

$$E\left(y_{t+2}|x_t = e_i\right) = \sum_k P_{ik}^{(2)}\overline{y}_k = \left(P^2\overline{y}\right)_i \tag{2.2.6}$$

and so on, where $P_{ik}^{(2)}$ denotes the (i,k) element of P^2. Notice that

$$\begin{aligned} E\left[E\left(y_{t+2}|x_{t+1} = e_j\right)|x_t = e_i\right] &= \sum_j P_{ij} \sum_k P_{jk}\overline{y}_k \\ &= \sum_k \left(\sum_j P_{ij}P_{jk}\right)\overline{y}_k = \sum_k P_{ik}^{(2)}\overline{y}_k = E\left(y_{t+2}|x_t = e_i\right). \end{aligned}$$

Connecting the first and last terms in this string of equalities yields $E[E(y_{t+2}|x_{t+1})|x_t] = E[y_{t+2}|x_t]$. This is an example of the 'law of iterated expectations'. The law of iterated expectations states that for any random variable z and two information sets J, I with $J \subset I$, $E[E(z|I)|J] = E(z|J)$. As another example of the law of iterated expectations, notice that

$$Ey_1 = \sum_j \pi_{1,j}\overline{y}_j = \pi_1'\overline{y} = \left(\pi_0' P\right)\overline{y} = \pi_0'\left(P\overline{y}\right)$$

and that

$$E\left[E\left(y_1|x_0 = e_i\right)\right] = \sum_i \pi_{0,i}\sum_j P_{ij}\overline{y}_j = \sum_j\left(\sum_i \pi_{0,i}P_{ij}\right)\overline{y}_j = \pi_1'\overline{y} = Ey_1.$$

2.2.4. Forecasting functions

There are powerful formulas for forecasting functions of a Markov process. Again let $\overline{y}$ be an $n \times 1$ vector and consider the random variable $y_t = \overline{y}' x_t$. Then

$$E\left[y_{t+k} | x_t = e_i\right] = \left(P^k \overline{y}\right)_i$$

where $(P^k \overline{y})_i$ denotes the ith row of $P^k \overline{y}$. Stacking all n rows together, we express this as

$$E\left[y_{t+k} | x_t\right] = P^k \overline{y}. \tag{2.2.7}$$

We also have

$$\sum_{k=0}^{\infty} \beta^k E\left[y_{t+k} | x_t = \overline{e}_i\right] = \left[\left(I - \beta P\right)^{-1} \overline{y}\right]_i ,$$

where $\beta \in (0,1)$ guarantees existence of $(I - \beta P)^{-1} = (I + \beta P + \beta^2 P^2 + \cdots)$.

One-step-ahead forecasts of a sufficiently rich set of random variables characterize a Markov chain. In particular, one-step-ahead conditional expectations of n independent functions (i.e., n linearly independent vectors $h_1, \ldots, h_n$) uniquely determine the transition matrix P. Thus, let $E[h_{k,t+1} | x_t = e_i] = (P h_k)_i$. We can collect the conditional expectations of h_k for all initial states i in an $n \times 1$ vector $E[h_{k,t+1} | x_t] = P h_k$. We can then collect conditional expectations for the n independent vectors $h_1, \ldots, h_n$ as $Ph = J$ where $h = [\, h_1 \quad h_2 \quad \ldots \quad h_n \,]$ and J is an the $n \times n$ matrix of all conditional expectations of all n vectors $h_1, \ldots, h_n$. If we know h and J, we can determine P from $P = J h^{-1}$.

2.2.5. Invariant functions and ergodicity

Let P, π be a stationary n-state Markov chain with the same state space we have chosen above, namely, $X = [e_i, i = 1, \ldots, n]$. An $n \times 1$ vector $\overline{y}$ defines a random variable $y_t = \overline{y}' x_t$. Thus, a random variable is another term for 'function of the underlying Markov state'.

The following is a useful precursor to a law of large numbers:

Theorem 2.2.1. *Let $\overline{y}$ define a random variable as a function of an underlying state x, where x is governed by a stationary Markov chain (P, π). Then*

$$\frac{1}{T} \sum_{t=1}^{T} y_t \to E\left[y_\infty | x_0\right] \tag{2.2.8}$$

with probability 1.

Here $E[y_\infty|x_0]$ is the expectation of y_s for s very large, conditional on the initial state. We want more than this. In particular, we would like to be able to replace $E[y_\infty|x_0]$ with the constant unconditional mean $E[y_t] = E[y_0]$ associated with the stationary distribution. To get this requires that we strengthen what is assumed about P by using the following concepts. First, we use

Definition 2.2.1. A random variable $y_t = \overline{y}' x_t$ is said to be *invariant* if $y_t = y_0, t \geq 0$, for any realization of $x_t, t \geq 0$.

Thus, a random variable y is invariant (or 'an invariant function of the state') if it remains constant while the underlying state x_t moves through the state space X.

For a finite state Markov chain, the following theorem gives a convenient way to characterize invariant functions of the state.

Theorem 2.2.2. *Let* P, π *be a stationary Markov chain. If*

$$E\left[y_{t+1}|x_t\right] = y_t \tag{2.2.9}$$

then the random variable $y_t = \overline{y}' x_t$ *is invariant.*

Proof. By using the law of iterated expectations, notice that

$$\begin{aligned} E\left(y_{t+1} - y_t\right)^2 &= E\left[E\left(y_{t+1}^2 - 2y_{t+1}y_t + y_t^2\right)|x_t\right] \\ &= E\left[Ey_{t+1}^2|x_t - 2E\left(y_{t+1}|x_t\right)y_t + Ey_t^2|x_t\right] \\ &= Ey_{t+1}^2 - 2Ey_t^2 + Ey_t^2 \\ &= 0 \end{aligned}$$

where the middle term in the right side of the second line uses that $E[y_t|x_t] = y_t$, the middle term on the right side of the third line uses the hypothesis (2.2.9), and the third line uses the hypothesis that π is a stationary distribution. In a finite Markov chain, if $E(y_{t+1} - y_t)^2 = 0$, then $y_{t+1} = y_t$ for all y_{t+1}, y_t that occur with positive probability under the stationary distribution. ∎

As we shall have reason to study in chapters 16 and 17, *any* (non necessarily stationary) stochastic process y_t that satisfies (2.2.9) is said to be a *martingale*. Theorem 2.2.2 tells us that a martingale that is a function of a finite state stationary Markov state x_t must be constant over time. This result is a special case of the martingale convergence theorem that underlies some remarkable results about savings to be studied in chapter 16.[1]

Equation (2.2.9) can be expressed as $P\overline{y} = \overline{y}$ or

$$(P - I)\,\overline{y} = 0, \tag{2.2.10}$$

which states that an invariant function of the state is a (right) eigenvector of P associated with a unit eigenvalue.

Definition 2.2.2. Let (P, π) be a stationary Markov chain. The chain is said to be *ergodic* if the only invariant functions $\overline{y}$ are constant with probability one, i.e., $\overline{y}_i = \overline{y}_j$ for all i, j with $\pi_i > 0, \pi_j > 0$.

A law of large numbers for Markov chains is:

Theorem 2.2.3. *Let $\overline{y}$ define a random variable on a stationary and ergodic Markov chain (P, π). Then*

$$\frac{1}{T}\sum_{t=1}^{T} y_t \to E\,[y_0] \tag{2.2.11}$$

with probability 1.

This theorem tells us that the time series average converges to the population mean of the stationary distribution.

Three examples illustrate these concepts.

Example 1. A chain with transition matrix $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has a unique invariant distribution $\pi = [\,.5 \quad .5\,]'$ and the invariant functions are $[\,\alpha \quad \alpha\,]'$ for any scalar α. Therefore the process is ergodic and Theorem 2.2.3 applies.

Example 2. A chain with transition matrix $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ has a continuum of stationary distributions $\gamma \begin{bmatrix} 1 \\ 0 \end{bmatrix} + (1-\gamma) \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ for any $\gamma \in [0,1]$ and invariant functions $\begin{bmatrix} 0 \\ \alpha \end{bmatrix}$ and $\begin{bmatrix} \alpha \\ 0 \end{bmatrix}$ for any α. Therefore, the process is not ergodic. The conclusion

[1] Theorem 2.2.2 tells us that a stationary martingale process has so little freedom to move that it has to be constant forever, not just eventually as asserted by the martingale convergence theorem.

(2.2.11) of Theorem 2.2.3 does not hold for many of the stationary distributions associated with P but Theorem 2.2.1 does hold. Conclusion (2.2.11) does hold for one particular choice of stationary distribution.

Example 3. A chain with transition matrix $P = \begin{bmatrix} .8 & .2 & 0 \\ .1 & .9 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ has a continuum of stationary distributions $\gamma\,[\,\frac{1}{3} \quad \frac{2}{3} \quad 0\,]' + (1-\gamma)\,[\,0 \quad 0 \quad 1\,]'$ and invariant functions $\alpha\,[\,1 \quad 1 \quad 0\,]'$ and $\alpha\,[\,0 \quad 0 \quad 1\,]'$ for any scalar α. The conclusion (2.2.11) of Theorem 2.2.3 does not hold for many of the stationary distributions associated with P but Theorem 2.2.1 does hold. But again, conclusion (2.2.11) does hold for one particular choice of stationary distribution.

2.2.6. Simulating a Markov chain

It is easy to simulate a Markov chain using a random number generator. The Matlab program `markov.m` does the job. We'll use this program in some later chapters.[2]

2.2.7. The likelihood function

Let P be an $n \times n$ stochastic matrix with states $1, 2, \ldots, n$. Let π_0 be an $n \times 1$ vector with nonnegative elements summing to 1, with $\pi_{0,i}$ being the probability that the state is i at time 0. Let i_t index the state at time t. The Markov property implies that the probability of drawing the path $(x_0, x_1, \ldots, x_{T-1}, x_T) = (\overline{e}_{i_0}, \overline{e}_{i_1}, \ldots, \overline{e}_{i_{T-1}}, \overline{e}_{i_T})$ is

$$\begin{aligned} L &\equiv \text{Prob}\left(\overline{x}_{i_T}, \overline{x}_{i_{T-1}}, \ldots, \overline{x}_{i_1}, \overline{x}_{i_0}\right) \\ &= P_{i_{T-1},i_T} P_{i_{T-2},i_{T-1}} \cdots P_{i_0,i_1} \pi_{0,i_0}. \end{aligned} \tag{2.2.12}$$

The probability L is called the *likelihood*. It is a function of both the sample realization $x_0, \ldots, x_T$ and the parameters of the stochastic matrix P. For a sample $x_0, x_1, \ldots, x_T$, let n_{ij} be the number of times that there occurs a one-period transition from state i to state j. Then the likelihood function can be written

$$L = \pi_{0,i_0} \prod_i \prod_j P_{i,j}^{n_{ij}},$$

[2] An index in the back of the book lists Matlab programs that can downloaded from the textbook web site <ftp://zia.stanford.edu/~sargent/pub/webdocs/matlab>.

a *multinomial* distribution.

Formula (2.2.12) has two uses. A first, which we shall encounter often, is to describe the probability of alternative histories of a Markov chain. In chapter 8, we shall use this formula to study prices and allocations in competitive equilibria.

A second use is for estimating the parameters of a model whose solution is a Markov chain. Maximum likelihood estimation for free parameters θ of a Markov process works as follows. Let the transition matrix P and the initial distribution π_0 be functions $P(\theta), \pi_0(\theta)$ of a vector of free parameters θ. Given a sample $\{x_t\}_{t=0}^T$, regard the likelihood function as a function of the parameters θ. As the estimator of θ, choose the value that maximizes the likelihood function L.

2.3. Continuous state Markov chain

In chapter 8 we shall use a somewhat different notation to express the same ideas. This alternative notation can accommodate either discrete or continuous state Markov chains. We shall let S denote the state space with typical element $s \in S$. The *transition density* is $\pi(s'|s) = \text{Prob}(s_{t+1} = s'|s_t = s)$ and the initial density is $\pi_0(s) = \text{Prob}(s_0 = s)$. For all $s \in S, \pi(s'|s) \geq 0$ and $\int_{s'} \pi(s'|s)ds' = 1$; also $\int_s \pi_0(s)ds = 1$.[3] Corresponding to (2.2.12), the likelihood function or density over the history $s^t = [s_t, s_{t-1}, \ldots, s_0]$ is

$$\pi\left(s^t\right) = \pi\left(s_t|s_{t-1}\right) \cdots \pi\left(s_1|s_0\right) \pi_0\left(s_0\right). \tag{2.3.1}$$

For $t \geq 1$, the time t unconditional distributions evolve according to

$$\pi_t\left(s_t\right) = \int_{s_{t-1}} \pi\left(s_t|s_{t-1}\right) \pi_{t-1}\left(s_{t-1}\right) d\, s_{t-1}.$$

A stationary or *invariant* distribution satisfies

$$\pi_\infty\left(s'\right) = \int_s \pi\left(s'|s\right) \pi_\infty\left(s\right) d\, s$$

which is the counterpart to (2.2.3).

3 Thus, when S is discrete, $\pi(s_j|s_i)$ corresponds to P_{s_i,s_j} in our earlier notation.

Paralleling our discussion of finite state Markov chains, we can say that the function $\phi(s)$ is invariant if

$$\int \phi\left(s'\right) \pi\left(s'|s\right) ds' = \phi\left(s\right).$$

A stationary continuous state Markov process is said to be *ergodic* if the only invariant functions $p(s')$ are constant with probability one according to the stationary distribution π_∞. A law of large numbers for Markov processes states:

Theorem 2.3.1. *Let $y(s)$ be a random variable, a measurable function of s, and let $(\pi(s'|s), \pi_0(s))$ be a stationary and ergodic continuous state Markov process. Assume that $E|y| < +\infty$. Then*

$$\frac{1}{T}\sum_{t=1}^{T} y_t \to Ey = \int y\left(s\right)\pi_0\left(s\right) ds$$

with probability 1 with respect to the distribution π_0.

2.4. Stochastic linear difference equations

The first order linear vector stochastic difference equation is a useful example of a continuous state Markov process. Here we could use $x_t \in I\!R^n$ rather than s_t to denote the time t state and specify that the initial distribution $\pi_0(x_0)$ is Gaussian with mean μ_0 and covariance matrix Σ_0; and that the transition density $\pi(x'|x)$ is Gaussian with mean $A_o x$ and covariance CC'. This specification pins down the joint distribution of the stochastic process $\{x_t\}_{t=0}^{\infty}$ via formula (2.3.1). The joint distribution determines all of the moments of the process that exist.

This specification can be represented in terms of the first-order stochastic linear difference equation

$$x_{t+1} = A_o x_t + C w_{t+1} \tag{2.4.1}$$

for $t = 0, 1, \ldots$, where x_t is an $n \times 1$ state vector, x_0 is a given initial condition, A_o is an $n \times n$ matrix, C is an $n \times m$ matrix, and w_{t+1} is an $m \times 1$ vector satisfying the following:

Assumption A1: w_{t+1} is an i.i.d. process satisfying $w_{t+1} \sim \mathcal{N}(0, I)$.

We can weaken the Gaussian assumption A1. To focus only on first and second moments of the x process, it is sufficient to make the weaker assumption:

ASSUMPTION A2: w_{t+1} is an $m \times 1$ random vector satisfying:

$$Ew_{t+1}|J_t = 0 \tag{2.4.2a}$$

$$Ew_{t+1}w'_{t+1}|J_t = I, \tag{2.4.2b}$$

where $J_t = [\,w_t \quad \cdots \quad w_1 \quad x_0\,]$ is the information set at t, and $E[\,\cdot\,|J_t]$ denotes the conditional expectation. We impose no distributional assumptions beyond (2.4.2). A sequence $\{w_{t+1}\}$ satisfying equation (2.4.2a) is said to be a martingale difference sequence adapted to J_t. A sequence $\{z_{t+1}\}$ that satisfies $E[z_{t+1}|J_t] = z_t$ is said to be a martingale adapted to J_t.

An even weaker assumption is

ASSUMPTION A3: w_{t+1} is a process satisfying

$$Ew_{t+1} = 0$$

for all t and

$$Ew_t w'_{t-j} = \begin{cases} I, & \text{if } j = 0; \\ 0, & \text{if } j \neq 0. \end{cases}$$

A process satisfying Assumption A3 is said to be a vector 'white noise'.[4]

Assumption A1 or A2 implies assumption A3 but not vice versa. Assumption A1 implies assumption A2 but not vice versa. Assumption A3 is sufficient to justify the formulas that we report below for second moments. We shall often append an observation equation $y_t = Gx_t$ to equation (2.4.1) and deal with the augmented system

$$x_{t+1} = A_o x_t + Cw_{t+1} \tag{2.4.3a}$$

$$y_t = Gx_t. \tag{2.4.3b}$$

Here y_t is a vector of variables observed at t, which may include only some linear combinations of x_t. The system (2.4.3) is often called a linear *state-space system.*

[4] Note that (2.4.2a) allows the distribution of w_{t+1} conditional on J_t to be heteroskedastic.

Example 1. Scalar second-order autoregression: Assume that z_t and w_t are scalar processes and that

$$z_{t+1} = \alpha + \rho_1 z_t + \rho_2 z_{t-1} + w_{t+1}.$$

Represent this relationship as the system

$$\begin{bmatrix} z_{t+1} \\ z_t \\ 1 \end{bmatrix} = \begin{bmatrix} \rho_1 & \rho_2 & \alpha \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_t \\ z_{t-1} \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} w_{t+1}$$
$$z_t = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_t \\ z_{t-1} \\ 1 \end{bmatrix}$$

which has form (2.4.3).

Example 2. First-order scalar mixed moving average and autoregression: Let

$$z_{t+1} = \rho z_t + w_{t+1} + \gamma w_t.$$

Express this relationship as

$$\begin{bmatrix} z_{t+1} \\ w_{t+1} \end{bmatrix} = \begin{bmatrix} \rho & \gamma \\ 0 & 0 \end{bmatrix} \begin{bmatrix} z_t \\ w_t \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} w_{t+1}$$
$$z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} z_t \\ w_t \end{bmatrix}.$$

Example 3. Vector autoregression: Let z_t be an $n \times 1$ vector of random variables. We define a vector autoregression by a stochastic difference equation

$$z_{t+1} = \sum_{j=1}^{4} A_j z_{t+1-j} + C_y w_{t+1}, \tag{2.4.4}$$

where w_{t+1} is an $n \times 1$ martingale difference sequence satisfying equation (2.4.2) with $x_0' = \begin{bmatrix} z_0 & z_{-1} & z_{-2} & z_{-3} \end{bmatrix}$ and A_j is an $n \times n$ matrix for each j. We can map equation (2.4.4) into equation (2.4.1) as follows:

$$\begin{bmatrix} z_{t+1} \\ z_t \\ z_{t-1} \\ z_{t-2} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & A_3 & A_4 \\ I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \end{bmatrix} \begin{bmatrix} z_t \\ z_{t-1} \\ z_{t-2} \\ z_{t-3} \end{bmatrix} + \begin{bmatrix} C_y \\ 0 \\ 0 \\ 0 \end{bmatrix} w_{t+1}. \tag{2.4.5}$$

Define A_o as the state transition matrix in equation (2.4.5). Assume that A_o has all of its eigenvalues bounded in modulus below unity. Then equation (2.4.4) can be initialized so that z_t is "covariance stationary," a term we now define.

2.4.1. First and second moments

We can use equation (2.4.1) to deduce the first and second moments of the sequence of random vectors $\{x_t\}_{t=0}^{\infty}$. A sequence of random vectors is called a stochastic process.

DEFINITION: A stochastic process $\{x_t\}$ is said to be *covariance stationary* if it satisfies the following two properties: (a) the mean is independent of time, $Ex_t = Ex_0$ for all t, and (b) the sequence of autocovariance matrices $E(x_{t+j} - Ex_{t+j})(x_t - Ex_t)'$ depends on the separation between dates $j = 0, \pm 1, \pm 2, \ldots$, but not on t.

We use

Definition 2.4.1. A square real valued matrix A is said to be *stable* if all of its eigenvalues have real parts that are strictly less than unity.

We shall often find it useful to assume that (2.4.3) takes the special form

$$\begin{bmatrix} x_{1,t+1} \\ x_{2,t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \tilde{A} \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2t} \end{bmatrix} + \begin{bmatrix} 0 \\ \tilde{C} \end{bmatrix} w_{t+1} \tag{2.4.6}$$

where $\tilde{A}$ is a stable matrix. That $\tilde{A}$ is a stable matrix implies that the only solution of $(\tilde{A} - I)\mu_2 = 0$ is $\mu_2 = 0$ (i.e., 1 is *not* an eigenvalue of $\tilde{A}$). It follows that the matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & \tilde{A} \end{bmatrix}$ on the right side of (2.4.6) has one eigenvector associated with a single unit eigenvalue: $(A - I)\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = 0$ implies μ_1 is an arbitrary scalar and $\mu_2 = 0$. The first equation of (2.4.6) implies that $x_{1,t+1} = x_{1,0}$ for all $t \geq 0$. Picking the initial condition $x_{1,0}$ pins down a particular eigenvector $\begin{bmatrix} x_{1,0} \\ 0 \end{bmatrix}$ of A. As we shall see soon, this eigenvector is our candidate for the unconditional mean of x that makes the process covariance stationary.

We will make an assumption that guarantees that there exists an initial condition $(Ex_0, E(x - Ex_0)(x - Ex_0)')$ that makes the x_t process covariance stationary. Either of the following conditions works:

CONDITION A1: All of the eigenvalues of A in (2.4.3) are strictly less than one in modulus.

CONDITION A2: The state space representation takes the special form (2.4.6) and all of the eigenvalues of $\tilde{A}$ are strictly less than one in modulus.

To discover the first and second moments of the x_t process, we regard the initial condition x_0 as being drawn from a distribution with mean $\mu_0 = Ex_0$ and covariance $\Sigma_0 = E(x - Ex_0)(x - Ex_0)'$. We shall deduce starting values for the mean and covariance that make the process covariance stationary, though our formulas are also useful for describing what happens when we start from some initial conditions that generate transient behavior that stops the process from being covariance stationary.

Taking mathematical expectations on both sides of equation (2.4.1) gives

$$\mu_{t+1} = A_o \mu_t \tag{2.4.7}$$

where $\mu_t = Ex_t$. We will assume that all of the eigenvalues of A_o are strictly less than unity in modulus, except possibly for one that is affiliated with the constant terms in the various equations. Then x_t possesses a stationary mean defined to satisfy $\mu_{t+1} = \mu_t$, which from equation (2.4.7) evidently satisfies

$$(I - A_o)\,\mu = 0, \tag{2.4.8}$$

which characterizes the mean μ as an eigenvector associated with the single unit eigenvalue of A_o. Notice that

$$x_{t+1} - \mu_{t+1} = A_o\,(x_t - \mu_t) + C w_{t+1}. \tag{2.4.9}$$

Also, the fact that the remaining eigenvalues of A_o are less than unity in modulus implies that starting from any μ_0, $\mu_t \to \mu$.[5]

From equation (2.4.9) we can compute that the stationary variance matrix satisfies

$$E\,(x_{t+1} - \mu)\,(x_{t+1} - \mu)' = A_o E\,(x_t - \mu)\,(x_t - \mu)'\,A_o' + CC'$$

or

$$C_x\,(0) \equiv E\,(x_t - \mu)\,(x_t - \mu)' = A_o C_x\,(0)\,A_o' + CC'. \tag{2.4.10}$$

[5] To see this point, assume that the eigenvalues of A_o are distinct, and use the representation $A_o = P\Lambda P^{-1}$ where Λ is a diagonal matrix of the eigenvalues of A_o, arranged in descending order in magnitude, and P is a matrix composed of the corresponding eigenvectors. Then equation (2.4.7) can be represented as $\mu^*_{t+1} = \Lambda \mu^*_t$, where $\mu^*_t \equiv P^{-1}\mu_t$, which implies that $\mu^*_t = \Lambda^t \mu^*_0$. When all eigenvalues but the first are less than unity, Λ^t converges to a matrix of zeros except for the $(1,1)$ element, and μ^*_t converges to a vector of zeros except for the first element, which stays at $\mu^*_{0,1}$, its initial value, which equals 1, to capture the constant. Then $\mu_t = P\mu^*_t$ converges to $P_1 \mu^*_{0,1} = P_1$, where P_1 is the eigenvector corresponding to the unit eigenvalue.

By virtue of (2.4.1) and (2.4.7), note that

$$(x_{t+j} - \mu_{t+j}) = A_o^j (x_t - \mu_t) + C w_{t+j} + \cdots + A_o^{j-1} C w_{t+1}.$$

Postmultiplying both sides by $(x_t - \mu_t)'$ and taking expectations shows that the autocovariance sequence satisfies

$$C_x (j) \equiv E (x_{t+j} - \mu) (x_t - \mu)' = A_o^j C_x (0) . \tag{2.4.11}$$

The autocovariance sequence is also called the *autocovariogram.* Equation (2.4.10) is a *discrete Lyapunov* equation in the $n \times n$ matrix $C_x(0)$. It can be solved with the Matlab program `doublej.m`. Once it is solved, the remaining second moments $C_x(j)$ can be deduced from equation (2.4.11).[6]

Suppose that $y_t = G x_t$. Then $\mu_{yt} = E y_t = G \mu_t$ and

$$E (y_{t+j} - \mu_{yt+j}) (y_t - \mu_{yt})' = G C_x (j) G', \tag{2.4.12}$$

for $j = 0, 1, \ldots$. Equations (2.4.12) are matrix versions of the so-called Yule-Walker equations, according to which the autocovariogram for a stochastic process governed by a stochastic linear difference equation obeys the nonstochastic version of that difference equation.

2.4.2. Impulse response function

Suppose that the eigenvalues of A_o not associated with the constant are bounded above in modulus by unity. Using the lag operator L defined by $L x_{t+1} \equiv x_t$, express equation (2.4.1) as

$$(I - A_o L) x_{t+1} = C w_{t+1}. \tag{2.4.13}$$

Recall the Neumann expansion $(I - A_o L)^{-1} = (I + A_o L + A_o^2 L^2 + \cdots)$ and apply $(I - A_o L)^{-1}$ to both sides of equation (2.4.13) to get

$$x_{t+1} = \sum_{j=0}^{\infty} A_o^j C w_{t+1-j}, \tag{2.4.14}$$

6 Notice that $C_x(-j) = C_x(j)'$.

which is the solution of equation (2.4.1) assuming that equation (2.4.1) has been operating for the infinite past before $t = 0$. Alternatively, iterate equation (2.4.1) forward from $t = 0$ to get

$$x_t = A_o^t x_0 + \sum_{j=0}^{t-1} A_o^j C w_{t-j} \tag{2.4.15}$$

Evidently,

$$y_t = G A_o^t x_0 + G \sum_{j=0}^{t-1} A_o^j C w_{t-j} \tag{2.4.16}$$

Equations (2.4.14), (2.4.15), and (2.4.16) are alternative versions of a moving average representation. Viewed as a function of lag j, $h_j = A_o^j C$ or $\tilde{h}_j = G A_o^j C$ is called the impulse response function. The moving average representation and the associated impulse response function show how x_{t+1} or y_{t+j} is affected by lagged values of the shocks, the w_{t+1}'s. Thus, the contribution of a shock w_{t-j} to x_t is $A_o^j C$.[7]

2.4.3. Prediction and discounting

From equation (2.4.1) we can compute the useful prediction formulas

$$E_t x_{t+j} = A_o^j x_t \tag{2.4.17}$$

for $j \geq 1$, where $E_t(\cdot)$ denotes the mathematical expectation conditioned on $x^t = (x_t, x_{t-1}, \ldots, x_0)$. Let $y_t = G x_t$, and suppose that we want to compute $E_t \sum_{j=0}^{\infty} \beta^j y_{t+j}$. Evidently,

$$E_t \sum_{j=0}^{\infty} \beta^j y_{t+j} = G \left(I - \beta A_o\right)^{-1} x_t, \tag{2.4.18}$$

provided that the eigenvalues of βA_o are less than unity in modulus. Equation (2.4.18) tells us how to compute an expected discounted sum, where the discount factor β is constant.

[7] The Matlab programs `dimpulse.m` and `impulse.m` compute impulse response functions.

2.4.4. Geometric sums of quadratic forms

In some applications, we want to calculate

$$\alpha_t = E_t \sum_{j=0}^{\infty} \beta^j x'_{t+j} Y x_{t+j}$$

where x_t obeys the stochastic difference equation (2.4.1) and Y is an $n \times n$ matrix. To get a formula for α_t, we use a guess-and-verify method. We guess that α_t can be written in the form

$$\alpha_t = x'_t \nu x_t + \sigma, \tag{2.4.19}$$

where ν is an $(n \times n)$ matrix, and σ is a scalar. The definition of α_t and the guess (2.4.19) imply

$$\begin{aligned}\alpha_t &= x'_t Y x_t + \beta E_t \left(x'_{t+1} \nu x_{t+1} + \sigma\right) \\ &= x'_t Y x_t + \beta E_t \left[(A_o x_t + C w_{t+1})' \ \nu \ (A_o x_t + C w_{t+1}) + \sigma\right] \\ &= x'_t \left(Y + \beta A'_o \nu A_o\right) x_t + \beta \ \text{trace}\,(\nu C C') + \beta \sigma.\end{aligned}$$

It follows that ν and σ satisfy

$$\begin{aligned}\nu &= Y + \beta A'_o \nu A_o \\ \sigma &= \beta \sigma + \beta \ \text{trace} \ \nu C C'.\end{aligned} \tag{2.4.20}$$

The first equation of (2.4.20) is a *discrete Lyapunov equation* in the square matrix ν, and can be solved by using one of several algorithms.[8] After ν has been computed, the second equation can be solved for the scalar σ.

We mention two important applications of formulas (2.4.19), (2.4.20).

Asset pricing

Let y_t be governed be governed by the state-space system (2.4.3). In addition, assume that there is a scalar random process z_t given by

$$z_t = H x_t.$$

[8] The Matlab control toolkit has a program called `dlyap.m` that works when all of the eigenvalues of A_o are strictly less than unity; the program called `doublej.m` works even when there is a unit eigenvalue associated with the constant.

Regard the process y_t as a payout or dividend from an asset, and regard $\beta^t z_t$ as a stochastic discount factor. The price of a perpetual claim on the stream of payouts is

$$\alpha_t = E_t \sum_{j=0}^{\infty} \left(\beta^j z_{t+j}\right) y_{t+j}. \tag{2.4.21}$$

To compute α_t, we simply set $Y = H'G$ in (2.4.19), (2.4.20). In this application, the term σ functions as a risk premium; it is zero when $C = 0$.

Evaluation of dynamic criterion

Let a state x_t be governed by

$$x_{t+1} = Ax_t + Bu_t + Cw_{t+1} \tag{2.4.22}$$

where u_t is a control vector that is set by a decision maker according to a fixed rule

$$u_t = -F_0 x_t. \tag{2.4.23}$$

Substituting (2.4.23) into (2.4.22) gives (2.4.1) where $A_o = A - BF_0$. We want to compute the *value function*

$$v(x_0) = -E_0 \sum_{t=0}^{\infty} \beta^t \left[x_t' R x_t + u_t' Q u_t\right]$$

for fixed matrices R and Q, fixed decision rule F_0 in (2.4.23), $A_o = A - BF_0$, and arbitrary initial condition x_0. Formulas (2.4.19), (2.4.20) apply with $Y = R + F_0' Q F_0$ and $A_o = A - BF_0$. Express the solution as

$$v(x_0) = -x_0' P x_0 - \sigma. \tag{2.4.24}$$

Now consider the following one-period problem. Suppose that we must use decision rule F_0 from time 1 onward, so that the value at time 1 on starting from state x_1 is

$$v(x_1) = -x_1' P x_1 - \sigma. \tag{2.4.25}$$

Taking $u_t = -F_0 x_t$ as given for $t \geq 1$, what is the best choice of u_0? This leads to the optimum problem:

$$\max_{u_0} -\{x_0' R x_0 + u_0' Q u_0 + \beta E \left(Ax_0 + Bu_0 + Cw_1\right)' P \left(Ax_0 + Bu_0 + Cw_1\right) + \beta\sigma\}. \tag{2.4.26}$$

The first-order conditions for this problem can be rearranged to attain

$$u_0 = -F_1 x_0 \tag{2.4.27}$$

where

$$F_1 = \beta \left(Q + \beta B' P B\right)^{-1} B' P A. \tag{2.4.28}$$

For convenience, we state the formula for P:

$$P = R + F_0' Q F_0 + \beta \left(A - B F_0\right)' P \left(A - B F_0\right). \tag{2.4.29}$$

Given F_0, formula (2.4.29) determines the matrix P in the value function that describes the expected discounted value of the sum of payoffs from sticking forever with this decision rule. Given P, formula (2.4.29) gives the best zero-period decision rule $u_0 = -F_1 x_0$ if you are permitted only a one-period deviation from the rule $u_t = -F_0 x_t$. If $F_1 \neq F_0$, we say that decision maker would accept the opportunity to deviate from F_0 for one period.

It is tempting to iterate on (2.4.28), (2.4.29) as follows to seek a decision rule from which a decision maker would not want to deviate for one period: (1) given an F_0, find P; (2) reset F equal to the F_1 found in step 1, then use (2.4.29) to compute a new P; (3) return to step 1 and iterate to convergence. This leads to the two equations

$$\begin{aligned} F_{j+1} &= \beta \left(Q + \beta B' P_j B\right)^{-1} B' P_j A \\ P_{j+1} &= R + F_j' Q F_j + \beta \left(A - B F_j\right)' P_{j+1} \left(A - B F_j\right). \end{aligned} \tag{2.4.30}$$

which are to be initialized from an arbitrary F_0 that assures that $\sqrt{\beta}(A - BF_0)$ is a stable matrix. After this process has converged, one cannot find a value-increasing one-period deviation from the limiting decision rule $u_t = -F_\infty x_t$.[9]

As we shall see in chapter 4, this is an excellent algorithm for solving a dynamic programming problem. It is called a Howard improvement algorithm.

[9] It turns out that if you don't want to deviate for one period, then you would never want to deviate, so that the limiting rule is optimal.

2.5. Population regression

This section explains the notion of a regression equation. Suppose that we have a state-space system (2.4.3) with initial conditions that make it covariance stationary. We can use the preceding formulas to compute the second moments of any pair of random variables. These moments let us compute a linear regression. Thus, let X be a $1 \times N$ vector of random variables somehow selected from the stochastic process $\{y_t\}$ governed by the system (2.4.3). For example, let $N = 2 \times m$, where y_t is an $m \times 1$ vector, and take $X = [\, y_t' \quad y_{t-1}' \,]$ for any $t \geq 1$. Let Y be any scalar random variable selected from the $m \times 1$ stochastic process $\{y_t\}$. For example, take $Y = y_{t+1,1}$ for the same t used to define X, where $y_{t+1,1}$ is the first component of y_{t+1}.

We consider the following least squares approximation problem: find an $N \times 1$ vector of real numbers β that attain

$$\min_{\beta} E\left(Y - X\beta\right)^2 \tag{2.5.1}$$

Here $X\beta$ is being used to estimate Y, and we want the value of β that minimizes the expected squared error. The first-order necessary condition for minimizing $E(Y - X\beta)^2$ with respect to β is

$$EX'\left(Y - X\beta\right) = 0, \tag{2.5.2}$$

which can be rearranged as $EX'Y = EX'X\beta$ or[10]

$$\beta = \left[E\left(X'X\right)\right]^{-1}\left(EX'Y\right). \tag{2.5.3}$$

By using the formulas (2.4.8), (2.4.10), (2.4.11), and (2.4.12), we can compute $EX'X$ and $EX'Y$ for whatever selection of X and Y we choose. The condition (2.5.2) is called the least squares normal equation. It states that the projection error $Y - X\beta$ is orthogonal to X. Therefore, we can represent Y as

$$Y = X\beta + \epsilon \tag{2.5.4}$$

where $EX'\epsilon = 0$. Equation (2.5.4) is called a regression equation, and $X\beta$ is called the least squares projection of Y on X or the least squares regression of Y on X. The vector β is called the population least squares regression vector. The law

[10] That $EX'X$ is nonnegative semidefinite implies that the second-order conditions for a minimum of condition (2.5.1) are satisfied.

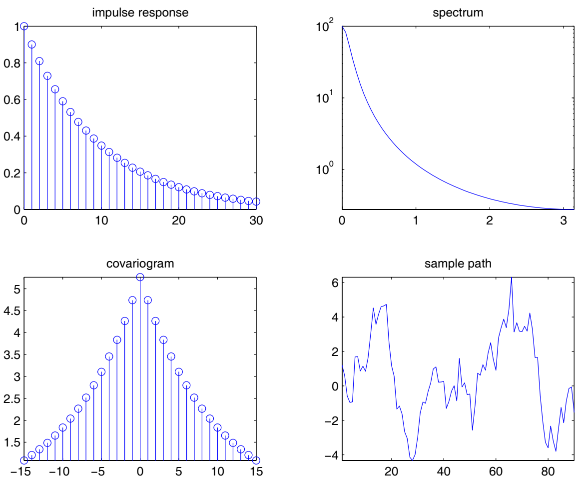

Figure 2.5.1: Impulse response, spectrum, covariogram, and sample path of process $(1 - .9L)y_t = w_t$.

of large numbers for continuous state Markov processes Theorem 2.3.1 states conditions that guarantee that sample moments converge to population moments, that is, $\frac{1}{S}\sum_{s=1}^{S} X_s' X_s \rightarrow EX'X$ and $\frac{1}{S}\sum_{s=1}^{S} X_s' Y_s \rightarrow EX'Y$. Under those conditions, sample least squares estimates converge to β.

There are as many such regressions as there are ways of selecting Y, X. We have shown how a model (e.g., a triple A_o, C, G, together with an initial distribution for x_0) restricts a regression. Going backward, that is, telling what a given regression tells about a model, is more difficult. Often the regression tells little about the model. The likelihood function encodes what a given data set says about the model.

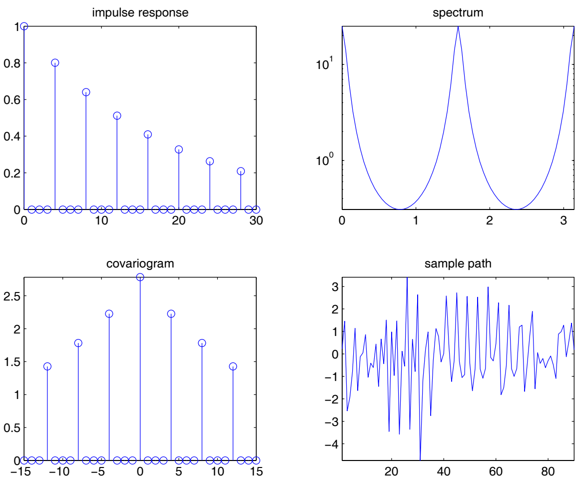

Figure 2.5.2: Impulse response, spectrum, covariogram, and sample path of process $(1 - .8L^4)y_t = w_t$.

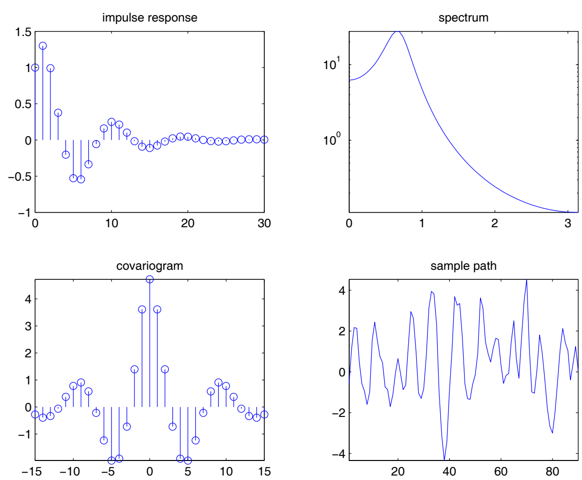

Figure 2.5.3: Impulse response, spectrum, covariogram, and sample path of process $(1 - 1.3L + .7L^2)y_t = w_t$.

2.5.1. The spectrum

For a covariance stationary stochastic process, all second moments can be encoded in a complex-valued matrix called the *spectral density* matrix. The autocovariance

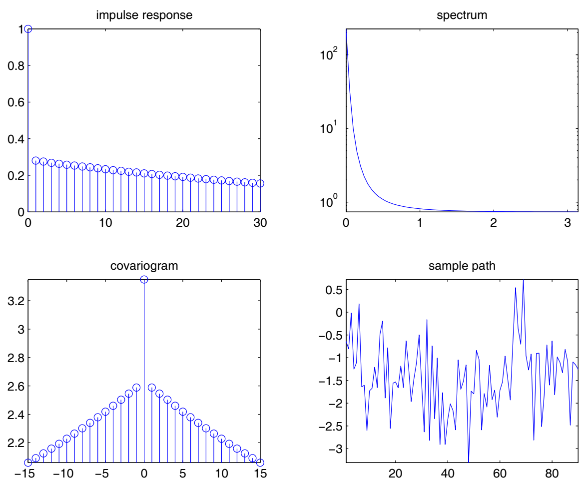

Figure 2.5.4: Impulse response, spectrum, covariogram, and sample path of process $(1 - .98L)y_t = (1 - .7L)w_t$.

sequence for the process determines the spectral density. Conversely, the spectral density can be used to determine the autocovariance sequence.

Under the assumption that A_o is a stable matrix,[11] the state x_t converges to a unique covariance stationary probability distribution as t approaches infinity. The spectral density matrix of this covariance stationary distribution $S_x(\omega)$ is defined to be the Fourier transform of the covariogram of x_t:

$$S_x(\omega) \equiv \sum_{\tau=-\infty}^{\infty} C_x(\tau) e^{-i\omega\tau}. \tag{2.5.5}$$

For the system (2.4.1), the spectral density of the stationary distribution is given by the formula

$$S_x(\omega) = \left[I - A_o e^{-i\omega}\right]^{-1} CC' \left[I - A_o' e^{+i\omega}\right]^{-1}, \quad \forall \omega \in [-\pi, \pi]. \tag{2.5.6}$$

The spectral density contains all of the information about the covariances. They can be recovered from $S_x(\omega)$ by the Fourier inversion formula[12]

11 It is sufficient that the only eigenvalue of A_o not strictly less than unity in modulus is that associated with the constant, which implies that A_o and C fit together in a way that validates (2.5.6).

12 Spectral densities for continuous-time systems are discussed by Kwakernaak and Sivan (1972). For an elementary discussion of discrete-time systems, see Sargent

$$C_x(\tau) = (1/2\pi) \int_{-\pi}^{\pi} S_x(\omega) e^{+i\omega\tau} d\omega.$$

Setting $\tau = 0$ in the inversion formula gives

$$C_x(0) = (1/2\pi) \int_{-\pi}^{\pi} S_x(\omega) d\omega,$$

which shows that the spectral density decomposes covariance across frequencies.[13] A formula used in the process of generalized method of moments (GMM) estimation emerges by setting $\omega = 0$ in equation (2.5.5), which gives

$$S_x(0) \equiv \sum_{\tau=-\infty}^{\infty} C_x(\tau).$$

2.5.2. Examples

To give some practice in reading spectral densities, we used the Matlab program `bigshow2.m` to generate Figures 2.5.1, 2.5.2, 2.5.4, and 2.5.3 The program takes as an input a univariate process of the form

$$a(L) y_t = b(L) w_t,$$

where w_t is a univariate martingale difference sequence with unit variance, where $a(L) = 1 - a_2 L - a_3 L^2 - \cdots - a_n L^{n-1}$ and $b(L) = b_1 + b_2 L + \cdots + b_n L^{n-1}$, and where we require that $a(z) = 0$ imply that $|z| > 1$. The program computes and displays a realization of the process, the impulse response function from w to y, and the spectrum of y. By using this program, a reader can teach himself to read spectra and impulse response functions. Figure 2.5.1 is for the pure autoregressive process with $a(L) = 1 - .9L, b = 1$. The spectrum sweeps downward in what C.W.J. Granger (1966) called the "typical spectral shape" for an economic time series. Figure 2.5.2 sets $a = 1 - .8L^4, b = 1$. This is a process with a strong seasonal component. That the spectrum peaks at π and $\pi/2$ are telltale signs of a strong seasonal component. Figure 2.5.4 sets $a = 1 - 1.3L + .7L^2, b = 1$. This is a process that has a spectral peak and cycles in its covariogram.[14] Figure 2.5.3 sets a $= 1 - .98L, b = 1 - .7L$.

(1987a). Also see Sargent (1987a, chap. 11) for definitions of the spectral density function and methods of evaluating this integral.

13 More interestingly, the spectral density achieves a decomposition of covariance into components that are orthogonal across frequencies.

14 See Sargent (1987a) for a more extended discussion.

This is a version of a process studied by Muth (1960). After the first lag, the impulse response declines as $.99^j$, where j is the lag length.

2.6. Example: the LQ permanent income model

To illustrate several of the key ideas of this chapter, this section describes the linear-quadratic savings problem whose solution is a rational expectations version of the permanent income model of Friedman (1956) and Hall (1978). We use this model as a vehicle for illustrating impulse response functions, alternative notions of the 'state', the idea of 'cointegration', and an invariant subspace method.

The LQ permanent income model is a modification (and not quite a special case for reasons that will be apparent later) of the following 'savings problem' to be studied in chapter 16. A consumer has preferences over consumption streams that are ordered by the utility functional

$$E_0 \sum_{t=0}^{\infty} \beta^t u(c_t) \tag{2.6.1}$$

where E_t is the mathematical expectation conditioned on the consumer's time t information, c_t is time t consumption and $u(c)$ is a strictly concave one-period utility function and $\beta \in (0,1)$ is a discount factor. The consumer maximizes (2.6.1) by choosing a consumption, borrowing plan $\{c_t, b_{t+1}\}_{t=0}^{\infty}$ subject to the sequence of budget constraints

$$c_t + b_t = R^{-1} b_{t+1} + y_t \tag{2.6.2}$$

where y_t is an exogenous stationary endowment process, R is a constant gross risk-free interest rate, b_t is one-period risk-free debt maturing at t, and b_0 is a given initial condition. We shall assume that $R^{-1} = \beta$. For example, we might assume that the endowment process has the state-space representation

$$z_{t+1} = A_{22} z_t + C_2 w_{t+1} \tag{2.6.3a}$$

$$y_t = U_y z_t \tag{2.6.3b}$$

where w_{t+1} is an i.i.d. process with mean zero and identify contemporaneous covariance matrix, A_{22} is a matrix the modulus of whose maximum eigenvalue is less than unity, and U_y is a selection vector that identifies y with a particular linear combination of the z_t. We impose the following condition on the consumption, borrowing plan:

$$E_0 \sum_{t=0}^{\infty} \beta^t b_t^2 < +\infty. \tag{2.6.4}$$

This condition suffices to rule out 'Ponzi schemes'. The *state* vector confronting the household at t is $[\,b_t \quad z_t\,]'$, where b_t is his one-period debt that falls due at the beginning of period t and z_t contains all variables useful for forecasting his future endowment. We impose this condition to rule out an always-borrow scheme that would allow the household to enjoy bliss consumption forever. The rationale for imposing this condition is to make the solution of the problem resemble more closely the solution of problems to be studied in chapter 16 that impose non-negativity on the consumption path. The first-order condition for maximizing (2.6.1) subject to (2.6.2) is[15]

$$E_t u'\left(c_{t+1}\right) = u'\left(c_t\right). \tag{2.6.5}$$

For the rest of this section we assume the quadratic utility function $u(c_t) = -.5(c_t - \gamma)^2$, where γ is a bliss level of consumption. Then (2.6.5) implies

$$E_t c_{t+1} = c_t. \tag{2.6.6}$$

Along with the quadratic utility specification, we allow consumption c_t to be negative.[16]

To deduce the optimal decision rule, we have to solve the system of difference equations formed by (2.6.2) and (2.6.6) subject to the boundary condition (2.6.4). To accomplish this, solve (2.6.2) forward to get

$$b_t = \sum_{j=0}^{\infty} \beta^j \left(y_{t+j} - c_{t+j}\right). \tag{2.6.7}$$

Take conditional expectations on both sides and use (2.6.6) and the law of iterated expectations to deduce

$$b_t = \sum_{j=0}^{\infty} \beta^j E_t y_{t+j} - \frac{1}{1-\beta} c_t \tag{2.6.8}$$

---

15 We shall study how to derive this first-order condition in detail in later chapters.

16 That c_t can be negative explains why we impose condition (2.6.4) instead of an upper bound on the level of borrowing, such as the natural borrowing limit of chapters 8, 16, and 17.

or

$$c_t = (1-\beta)\left[\sum_{j=0}^{\infty}\beta^j E_t y_{t+j} - b_t\right]. \tag{2.6.9}$$

If we define the net rate of interest r by $\beta = \frac{1}{1+r}$, we can also express this equation as

$$c_t = \frac{r}{1+r}\left[\sum_{j=0}^{\infty}\beta^j E_t y_{t+j} - b_t\right]. \tag{2.6.10}$$

Equation (2.6.9) or (2.6.10) expresses consumption as equalling economic *income*, namely, a constant marginal propensity consume or interest factor $\frac{r}{1+r}$ times the sum of non-financial wealth $\sum_{j=0}^{\infty}\beta^j E_t y_{t+j}$ and financial wealth $-b_t$. Notice that (2.6.9) or (2.6.10) represents c_t as a function of the *state* $[b_t, z_t]$ confronting the household, where from (2.6.3) z_t contains the information useful for forecasting the endowment process that enters the conditional expectation E_t.

A revealing way of understanding the solution is to show that *after* the optimal decision rule has been obtained, there is a point of view that allows us to regard the state as being c_t together with z_t and to regard b_t as an 'outcome'. Following Hall (1978), this is a sharp way to summarize the implication of the LQ permanent income theory. We now proceed to transform the state vector in this way.

To represent the solution for b_t, substitute (2.6.9) into (2.6.2) and after rearranging obtain

$$b_{t+1} = b_t + \left(\beta^{-1}-1\right)\sum_{j=0}^{\infty}\beta^j E_t y_{t+j} - \beta^{-1} y_t. \tag{2.6.11}$$

Next shift (2.6.9) forward one period and eliminate b_{t+1} by using (2.6.2) to obtain

$$c_{t+1} = (1-\beta)\sum_{j=0}^{\infty} E_{t+1}\beta^j y_{t+j+1} - (1-\beta)\left[\beta^{-1}\left(c_t + b_t - y_t\right)\right].$$

If we add and subtract $\beta^{-1}(1-\beta)\sum_{j=0}^{\infty}\beta^j E_t y_{t+j}$ from the right side of the preceding equation and rearrange, we obtain

$$c_{t+1} = c_t + (1-\beta)\sum_{j=0}^{\infty}\beta^j\left(E_{t+1}y_{t+j+1} - E_t y_{t+j+1}\right). \tag{2.6.12}$$

The right side is the time $t+1$ innovation to the expected present value of the endowment process y. Suppose that the endowment process has the moving average representation[17]

$$y_{t+1} = d\,(L)\, w_{t+1} \tag{2.6.13}$$

where w_{t+1} is an i.i.d. vector process with $Ew_{t+1} = 0$ and contemporaneous covariance matrix $Ew_{t+1}w_{t+1}' = I$, $d(L) = \sum_{j=0}^{\infty} d_j L^j$, where L is the lag operator, and the household has an information set $w^t = [w_t, w_{t-1}, \ldots,]$ at time t. Then notice that

$$y_{t+j} - E_t y_{t+j} = d_0 w_{t+j} + d_1 w_{t+j-1} + \cdots + d_{j-1} w_{t+1}.$$

It follows that

$$E_{t+1} y_{t+j} - E_t y_{t+j} = d_{j-1} w_{t+1}. \tag{2.6.14}$$

Using (2.6.14) in (2.6.12) gives

$$c_{t+1} - c_t = (1 - \beta)\, d\,(\beta)\, w_{t+1}. \tag{2.6.15}$$

The object $d(\beta)$ is the present value of the moving average coefficients in the representation for the endowment process y_t.

After all of this work, we can represent the optimal decision rule for c_t, b_{t+1} in the form of the two equations (2.6.12), (2.6.8), which we repeat here for convenience:

$$c_{t+1} = c_t + (1 - \beta) \sum_{j=0}^{\infty} \beta^j \left(E_{t+1} y_{t+j+1} - E_t y_{t+j+1}\right) \tag{2.6.16}$$

$$b_t = \sum_{j=0}^{\infty} \beta^j E_t y_{t+j} - \frac{1}{1-\beta} c_t. \tag{2.6.17}$$

Recalling the form of the endowment process (2.6.3), we can compute

$$\begin{aligned} E_t \sum_{j=0}^{\infty} \beta^j z_{t+j} &= (I - \beta A_{22})^{-1} z_t \\ E_{t+1} \sum_{j=0}^{\infty} \beta^j z_{t+j+1} &= (I - \beta A_{22})^{-1} z_{t+1} \\ E_t \sum_{j=0}^{\infty} \beta^j z_{t+j+1} &= (I - \beta A_{22})^{-1} A_{22} z_t. \end{aligned}$$

17 Representation (2.6.3) implies that $d(L) = U_y (I - A_{22} L)^{-1} C_2$.

Substituting these formulas into (2.6.16), (2.6.17) and using (2.6.3a) gives the following representation for the consumer's optimum decision rule:

$$c_{t+1} = c_t + (1-\beta)\, U_y \,(I-\beta A_{22})^{-1}\, C_2 w_{t+1} \tag{2.6.18a}$$

$$b_t = U_y\,(I-\beta A_{22})^{-1}\, z_t - \frac{1}{1-\beta} c_t \tag{2.6.18b}$$

$$y_t = U_y z_t \tag{2.6.18c}$$

$$z_{t+1} = A_{22} z_t + C_2 w_{t+1} \tag{2.6.18d}$$

Representation (2.6.18) reveals several things about the optimal decision rule. (1) The *state* consists of the endogenous part c_t and the exogenous part z_t. These contain all of the relevant information for forecasting future c, y, b. Notice that financial assets b_t have disappeared as a component of the state because they are properly encoded in c_t. (2) According to (2.6.18), consumption is a random walk with innovation $(1-\beta)d(\beta)w_{t+1}$ as implied also by (2.6.15). This outcome confirms that the Euler equation (2.6.6) is built into the solution. That consumption is a random walk of course implies that it does not possess an asymptotic stationary distribution, at least so long as z_t exhibits perpetual random fluctuations, as it will generally under (2.6.3).[18] This feature is inherited partly from the assumption that $\beta R = 1$. (3) The impulse response function of c_t is a 'box': for all $j \geq 1$, the response of c_{t+j} to an increase in the innovation w_{t+1} is $(1-\beta)d(\beta) = (1-\beta)U_y(I-\beta A_{22})^{-1}C_2$. (4) Solution (2.6.18) reveals that the joint process c_t, b_t possesses the property that Granger and Engle (1987) called cointegration. In particular, *both* c_t and b_t are non-stationary because they have unit roots (see representation (2.6.11) for b_t), but there is a linear combination of c_t, b_t that *is* stationary provided that z_t is stationary. From (2.6.17), the linear combination is $(1-\beta)b_t + c_t$. Accordingly, Granger and Engle would call $[\,(1-\beta) \quad 1\,]$ a 'co-integrating vector' that when applied to the nonstationary vector process $[\,b_t \quad c_t\,]'$ yields a process that is asymptotically stationary. Equation (2.6.8) can be arranged to take the form

$$(1-\beta)\, b_t + c_t = E_t \sum_{j=0}^{\infty} \beta^j y_{t+j}, \tag{2.6.19}$$

which asserts that the 'co-integrating residual' on the left side equals the conditional expectation of the geometric sum of future incomes on the right.[19] Lettau, Martin

[18] The failure of consumption to converge will also occur in chapter 16 when we drop quadratic utility and assume that consumption must be nonnegative.

[19] See Campbell and Shiller (1988) and Lettau and Ludvigson (2001, 2004) for interesting applications of related ideas.

2.6.1. Invariant subspace approach

We can glean additional insights about the structure of the optimal decision rule by solving the decision problem in a mechanical but quite revealing way that easily generalizes to a host of problems, as we shall see later in chapter 5. We can represent the system consisting of the Euler equation (2.6.6), the budget constraint (2.6.2), and the description of the endowment process (2.6.3) as

$$\begin{bmatrix} \beta & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_{t+1} \\ z_{t+1} \\ c_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & -U_y & 1 \\ 0 & A_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_t \\ z_t \\ c_t \end{bmatrix} + \begin{bmatrix} 0 \\ C_2 \\ C_1 \end{bmatrix} w_{t+1} \tag{2.6.20}$$

where C_1 is an undetermined coefficient. Premultiply both sides by the inverse of the matrix on the left and write

$$\begin{bmatrix} b_{t+1} \\ z_{t+1} \\ c_{t+1} \end{bmatrix} = \tilde{A} \begin{bmatrix} b_t \\ z_t \\ c_t \end{bmatrix} + \tilde{C} w_{t+1}. \tag{2.6.21}$$

We want to find solutions of (2.6.21) that satisfy the no-explosion condition (2.6.4). We can do this by using machinery from chapter 5. The key idea is to discover what part of the vector $\begin{bmatrix} b_t & z_t & c_t \end{bmatrix}'$ is truly a *state* from the view of the decision maker, being inherited form the past, and what part is a 'co-state' or 'jump' variable that can adjust at t. For our problem b_t, z_t are truly components of the state, but c_t is free to adjust. The theory determines c_t at t as a function of the true state variables $[b_t, z_t]$. A powerful approach to determining this function is the following so-called invariant subspace method of chapter 5. Obtain the eigenvector decomposition of $\tilde{A}$:

$$\tilde{A} = V \Lambda V^{-1}$$

where Λ is a diagonal matrix consisting of the eigenvalues of $\tilde{A}$ and V is a matrix of the associated eigenvectors. Let $V^{-1} \equiv \begin{bmatrix} V^{11} & V^{12} \\ V^{21} & V^{22} \end{bmatrix}$. Then applying formula (5.5.11) of chapter 5 implies that if (2.6.4) is to hold, then the jump variable c_t must satisfy

$$c_t = -\left(V^{22}\right)^{-1} V^{21} \begin{bmatrix} b_t \\ z_t \end{bmatrix}. \tag{2.6.22}$$

Formula (2.6.22) gives the unique value of c_t that assures that (2.6.4) is satisfied, or in other words, that the state remains in the 'stabilizing subspace'. Notice that the variables on the right side of (2.6.22) conform with those called for by (2.6.10): $-b_t$ is there as a measure of financial wealth, and z_t is there because it includes all variables that are useful for forecasting the future endowments that occur in (2.6.10).

2.7. The term structure of interest rates

Asset prices encode investors' expectations about future payoffs. If we suppose that investors form their expectations using versions of our optimal forecasting formulas, we acquire a theory of asset prices. Here we use the term structure of interest rates as an example.

2.7.1. A stochastic discount factor

Let's start with just a little background in the theory of asset pricing. To begin with the simplest case, let $\{d_t\}_{t=0}^{\infty}$ be a stream of dividends. Let p_t be the price of a claim on what remains of the dividend stream from date $t+1$ on. The standard asset pricing model under certainty asserts that

$$p_t = \sum_{j=1}^{\infty} \left(\prod_{s=1}^{j} m_{t+s} \right) d_{t+j} \tag{2.7.1}$$

where m_{t+1} is a one-period factor for discounting dividends between t and $t+1$ and $\prod_{s=1}^{j} m_{t+j}$ is a j-period factor for discounting dividends between $t+j$ and t. A simple model assumes a constant discount factor $m_s = \beta$, which makes (2.7.1) become

$$p_t = \sum_{j=1}^{\infty} \beta^j d_{t+j}.$$

In chapter 13, we shall study generalizations of (2.7.1) that take the form

$$p_t = E_t \sum_{j=1}^{\infty} \left(\prod_{s=1}^{j} m_{t+s} \right) d_{t+j} \tag{2.7.2}$$

where m_{t+1} is a one-period *stochastic discount factor* for converting a time $t+1$ payoff into a time t value, and E_t is a mathematical expectation conditioned on time

t information. In this section, we use a version of formula (2.7.2) to illustrate the power of our formulas for solving linear stochastic difference equations.

We specify a dividend process in a special way that is designed to make p_t be the price of an n-period risk-free pure discount nominal bond: $d_{t+n} = 1, d_{t+j} = 0$ for $j \neq n$, where for a nominal bond '1' means one dollar. In this case, we add a subscript n to help us remember the period for the bond and (2.7.2) becomes

$$p_{nt} = E_t \left(\prod_{s=1}^{n} m_{t+s} \right) \tag{2.7.3}$$

We define the yield y_{nt} on an n-period bond by $p_{nt} = \exp(-n y_{nt})$ or

$$y_{nt} = -n^{-1} \log p_{nt}. \tag{2.7.4}$$

Thus, yields are linear in the logs of the corresponding bond prices. Bond yields are Gaussian when bond prices are log-normal (i.e., the log of bond prices are Gaussian) and this will be the outcome if we specify that the log of the discount factor m_{t+1} follows a Gaussian process.

2.7.2. The log normal bond pricing model

Here is the log-normal bond price model. A one-period stochastic discount factor at t is m_{t+1} and an n-period stochastic discount factor at t is $m_{t+1} m_{t+2} \cdots m_{t+n}$.[20] The logarithm of the one-period stochastic discount factor follows the stochastic process

$$\log m_{t+1} = -\delta - e_z z_{t+1} \tag{2.7.5a}$$

$$z_{t+1} = A_z z_t + C_z w_{t+1} \tag{2.7.5b}$$

where w_{t+1} is an i.i.d. Gaussian random vector with $E w_{t+1} = 0$, $E w_{t+1} w_{t+1}' = I$, and A_z is an $m \times m$ matrix all of whose eigenvalues are bounded by unity in modulus. Soon we shall describe the process for the log of the nominal stochastic discount factor that Backus and Zin (1994) used to emulate the term structure of nominal interest rates in the U.S. during the post WWII period. At time t, an n-period risk free

[20] Some authors use the notation $m_{t+j,t}$ to denote a j-period stochastic discount factor at time t. The transformation between that notation and ours is $m_{t+1,t} = m_{t+1}, \ldots, m_{t+j,t} = m_{t+1} \cdots m_{t+j}$.

nominal bond promises to pay one dollar for sure in period $t+n$. According to (2.7.3), the price at t of this bond is the conditional expectation of the product of the n-period stochastic discount factor times the unit payout.[21] Applying (2.7.4) to (2.7.3) gives

$$y_{nt} = -n^{-1} \log E_t \left[m_{t+1} \cdots m_{t+n} \right]. \tag{2.7.6}$$

To evaluate the right side of (2.7.6), we use the following property of log normal distributions:

LOG NORMAL DISTRIBUTION: If $\log m_{t+1} \sim \mathcal{N}(\mu, \sigma^2)$ (i.e., $\log m_{t+1}$ is Gaussian with mean μ and variance σ^2), then

$$\log E m_{t+1} = \mu + \frac{\sigma^2}{2}. \tag{2.7.7}$$

Applying this property to the conditional distribution of m_{t+1} induced by (2.7.5) gives

$$\log E_t m_{t+1} = -\delta - e_z A_z z_t + \frac{e_z C_z {C_z}' {e_z}'}{2}. \tag{2.7.8}$$

By iterating on (2.7.5), we can obtain the following expression that is useful for characterizing the conditional distribution of $\log (m_{t+1} \cdots m_{t+n})$:

$$\begin{aligned} -\left(\log\left(m_{t+1}\right) + \cdots \log\left(m_{t+n}\right)\right) = n\delta &+ e_z \left(A_z + {A_z}^2 + \cdots {A_z}^n\right) z_t \\ &+ e_z C_z w_{t+n} + e_z \left[C_z + A_z C_z\right] w_{t+n-1} \\ &+ \cdots + e_z \left[C_z + A_z C_z + \cdots + {A_z}^{n-1} C_z\right] w_{t+1} \end{aligned} \tag{2.7.9}$$

The distribution of $\log m_{t+1} + \cdots \log m_{t+n}$ conditional on z_t is thus $\mathcal{N}(\mu_{nt}, \sigma_n^2)$, where[22]

$$\mu_{nt} = -\left[n\delta + e_z \left(A_z + \cdots {A_z}^n\right) z_t\right] \tag{2.7.10a}$$
$$\sigma_1^2 = e_z C_z {C_z}' {e_z}' \tag{2.7.10b}$$
$$\sigma_n^2 = \sigma_{n-1}^2 + e_z \left[I + \cdots + {A_z}^{n-1}\right] C_z {C_z}' \left[I + \cdots + {A_z}^{n-1}\right]' {e_z}' \tag{2.7.10c}$$

---

21 That is, the price of the bond is the price of the payouts times their quantities added across states via the expectation operator

22 For the purpose of programming these formulas, it is useful to note that $(I + A_z + \cdots + {A_z}^{n-1}) = (I - A_z)^{-1}(I - {A_z}^n)$.

where the recursion (2.7.10c) holds for $n \geq 2$. Notice that the conditional means μ_{nt} vary over time but that the conditional covariances σ_n^2 are constant over time.[23] Applying (2.7.6) and formula (2.7.7) for the log of the expectation of a log normally distributed random variable gives the following formula for bond yields:

$$y_{nt} = \left(\delta - \frac{\sigma_n^2}{2 \times n}\right) + n^{-1} e_z \left(A_z + \cdots + A_z{}^n\right) z_t. \tag{2.7.11}$$

The vector $y_t = [y_{1t} \quad y_{2t} \quad \cdots \quad y_{nt}]'$ is called the term structure of nominal interest rates at time t. A specification known as the *expectations theory of the term structure* resembles but differs from (2.7.11). The expectations theory asserts that n period yields are averages of expected future values of one-period yields, which translates to

$$y_{nt} = \delta + n^{-1} e_z \left(A_z + \cdots + A_z{}^n\right) z_t \tag{2.7.12}$$

because evidently the conditional expectation $E_t y_{1t+j} = \delta + e_z A_z^j z_t$. The expectations theory (2.7.12) can be viewed as an approximation to the log-normal yield model (2.7.11) that neglects the contributions of the variance terms σ_n^2 to the constant terms.

Returning to the log-normal bond price model, we evidently have the following compact state space representation for the term structure of interest rates and its dependence on the law of motion for the stochastic discount factor:

$$X_{t+1} = A_o X_t + C w_{t+1} \tag{2.7.13a}$$

$$Y_t \equiv \begin{bmatrix} y_t \\ \log m_t \end{bmatrix} = G X_t \tag{2.7.13b}$$

where

$$X_t = \begin{bmatrix} 1 \\ z_t \end{bmatrix} \quad A_o = \begin{bmatrix} 1 & 0 \\ 0 & A_z \end{bmatrix} \quad C = \begin{bmatrix} 0 \\ C_z \end{bmatrix}$$

and

$$G = \begin{bmatrix} \delta - \frac{\sigma_1^2}{2} & e_z A_z \\ \delta - \frac{\sigma_2^2}{2 \times 2} & 2^{-1} e_z \left(A_z + A_z{}^2\right) \\ \vdots & \vdots \\ \delta - \frac{\sigma_n^2}{2 \times n} & n^{-1} e_z \left(A_z + \cdots + A_z{}^n\right) \\ -\delta & -e_z \end{bmatrix}.$$

[23] The celebrated *affine* term structure model generalizes the log-normal model by allowing σ_n^2 to depend on time by feeding back on parts of the state vector. See Ang and Piazzesi (2003) for recent estimates of an affine term structure model.

2.7.3. Slope of yield curve depends on serial correlation of $\log m_{t+1}$

From (2.7.13), it follows immediately that the unconditional mean of the term structure is

$$Ey_t' = [\,\delta - \sigma_1^2 \quad \cdots \quad \delta - \tfrac{\sigma_n^2}{2\times n}\,]' ,$$

so that the term structure on average rises with horizon only if σ_j^2/j falls as j increases. By interpreting our formulas for the σ_j^2's, it is possible to show that a term structure that on average *rises* with maturity implies that the log of the stochastic discount factor is *negatively* serially correlated. Thus, it can be verified from (2.7.9) that the term σ_j^2 in (2.7.10) and (2.7.11) satisfies

$$\sigma_j^2 = \mathrm{var}_t\,(\log m_{t+1} + \cdots + \log m_{t+j})$$

where var_t denotes a variance conditioned on time t information z_t. Notice, for example, that

$$\mathrm{var}_t\,(\log m_{t+1} + \log m_{t+2}) = \mathrm{var}_t\,(\log m_{t+1}) + \mathrm{var}_t\,(\log m_{t+2}) + 2\mathrm{cov}_t\,(\log m_{t+1}, \log m_{t+2}) \tag{2.7.14}$$

where cov_t is a conditional covariance. It can then be established that $\sigma_1^2 > \frac{\sigma_2^2}{2}$ can occur only if $\mathrm{cov}_t(\log m_{t+1}, \log m_{t+2}) < 0$. Thus, a yield curve that is upward sloping on average reveals that the log of the stochastic discount factor is negatively serially correlated. (See the spectrum of the log stochastic discount factor in Fig. 2.7.5 below.)

2.7.4. Backus and Zin's stochastic discount factor

For a specification of A_z, C_z, δ for which the eigenvalues of A_z are all less than unity, we can use the formulas presented above to compute moments of the stationary distribution EY_t, as well as the autocovariance function $\mathrm{Cov}_Y(\tau)$ and the impulse response function given in (2.4.15) or (2.4.16). For the term structure of nominal U.S. interest rates over much of the post WWII period, Backus and Zin (1994) provide us with an empirically plausible specification of A_z, C_z, e_z. In particular, they specify that $\log m_{t+1}$ is a stationary autoregressive moving average process

$$-\phi\,(L) \log m_{t+1} = \phi\,(1)\,\delta + \theta\,(L)\,\sigma w_{t+1}$$

where w_{t+1} is a scalar Gaussian white noise with $Ew_{t+1}^2 = 1$ and

$$\phi\,(L) = 1 - \phi_1 L - \phi_2 L^2 \tag{2.7.15a}$$

$$\theta\,(L) = 1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3. \tag{2.7.15b}$$

Backus and Zin specified parameter values for that imply that all of the zeros of both $\phi(L)$ and $\theta(L)$ *exceed* unity in modulus,[24] a condition that assures that the eigenvalues of A_o are all *less than* unity in modulus. Backus and Zin's specification can be captured by setting

$$z_t = [\,\log m_t \quad \log m_{t-1} \quad w_t \quad w_{t-1} \quad w_{t-2}\,]$$

and

$$A_z = \begin{bmatrix} \phi_1 & \phi_2 & \theta_1\sigma & \theta_2\sigma & \theta_3\sigma \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

and $C_z = [\,\sigma \quad 0 \quad 1 \quad 0 \quad 0\,]'$ where $\sigma > 0$ is the standard deviation of the innovation to $\log m_{t+1}$ and $e_z = [\,1 \quad 0 \quad 0 \quad 0 \quad 0\,]$.

2.7.5. Reverse engineering a stochastic discount factor

Backus and Zin use time series data on y_t together with the restrictions implied by the log normal bond pricing model and to deduce implications about the stochastic discount factor m_{t+1}. They call this procedure 'reverse engineering the yield curve', but what they really do is use time series observations on the *yield curve* to reverse engineer a *stochastic discount factor*. They used the generalized method of moments to estimate (some people say 'calibrate') the following values for monthly U.S. nominal interest rates on pure discount bonds: $\delta = .528, \sigma = 1.023$, $\theta(L) = 1 - 1.031448L + .073011L^2 + .000322L^3$, $\phi(L) = 1 - 1.031253L + .073191L^2$. Why do Backus and Zin carry along so many digits? To explain why, first notice that with these particular values $\frac{\theta(L)}{\phi(L)} \approx 1$, so that the log of the stochastic discount factor is well approximated by an i.i.d. process:

$$-\log m_{t+1} \approx \delta + \sigma w_{t+1}.$$

This means that fluctuations in the log stochastic discount factor are difficult to predict. Backus and Zin argue convincingly that to match observed features that are summarized by estimated first and second moments of the nominal term structure y_t process and for yields on other risky assets for the U.S. after World War II, it

[24] A complex variable z_0 is said to be a zero of $\phi(z)$ if $\phi(z_0) = 0$.

is important that $\theta(L), \phi(L)$ have two properties: (a) first, $\theta(L) \approx \phi(L)$ so that the stochastic discount factor is volatile variable whose fluctuations are difficult to predict variable; and (b) nevertheless that $\theta(L) \neq \phi(L)$ so that the stochastic discount factor has subtle predictable components. Feature (a) is needed to match observed prices of risky securities, as we shall discuss in chapter 13. In particular, observations on returns on risky securities can be used to calculate a so-called 'market price of risk' that in theory should equal $\frac{\sigma_t(m_{t+1})}{E_t m_{t+1}}$, where σ_t denotes a conditional standard deviation and E_t a conditional mean, conditioned on time t information. Empirical estimates of the stochastic discount factor from the yield curve and other asset returns suggest a value of the market price of risk that is relatively large, in a sense that we explore in depth in chapter 13. A high volatility of m_{t+1} delivers a high market price of risk. Backus and Zin use feature (b) to match the shape of the yield curve over time. Backus and Zin's estimates of $\phi(L), \theta(L)$ imply term structure outcomes that display both features (a) and (b). For their values of $\theta(L), \phi(L), \sigma$, Fig. 2.7.1–Fig. 2.7.5 show various aspects of the theoretical yield curve. Fig. 2.7.1 shows the theoretical value of the mean term structure of interest rates, which we have calculated by applying our formula for $\mu_Y = G\mu_X$ to (2.7.13). The theoretical value of the yield curve is on average upward sloping, as is true also in the data. For yields of durations $j = 1, 3, 6, 12, 24, 36, 48, 60, 120, 360$, where duration is measured in *months*, Fig. 2.7.2 shows the impulse response of y_{jt} to a shock w_{t+1} in the log of the stochastic discount factor. We use formula (2.4.16) to compute this impulse response function. In Fig. 2.7.2, bigger impulse response functions are associated with *shorter* horizons. The shape of the impulse response function for the short rate differs from the others: it is the only one with a 'humped' shape. Fig. 2.7.3 and Fig. 2.7.4 show the impulse response function of the log of the stochastic discount factor. Fig. 2.7.3 confirms that $\log m_{t+1}$ is approximately i.i.d. (the impulse response occurs mostly at zero lag), but Fig. 2.7.4 shows the impulse response coefficients for lags of 1 and greater and confirms that the stochastic discount factor is not quite i.i.d. Since the initial response is a large negative number, these small positive responses for positive lags impart *negative* serial correlation to the log stochastic discount factor. As noted above and as stressed by Backus and Zin (1992), negative serial correlation of the stochastic discount factor is needed to account for a yield curve that is upward sloping on average.

Fig. 2.7.5 applies the Matlab program `bigshow2` to Backus and Zin's specified values of $(\sigma, \delta, \theta(L), \phi(L))$. The panel on the upper left is the impulse response again. The panel on the lower left shows the covariogram, which as expected is very close to that for an i.i.d. process. The spectrum of the log stochastic discount factor is

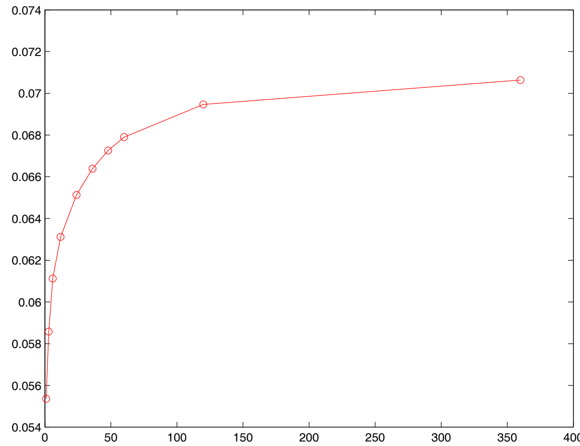

Figure 2.7.1: Mean term structure of interest rates with Backus-Zin stochastic discount factor (months on horizontal axis).

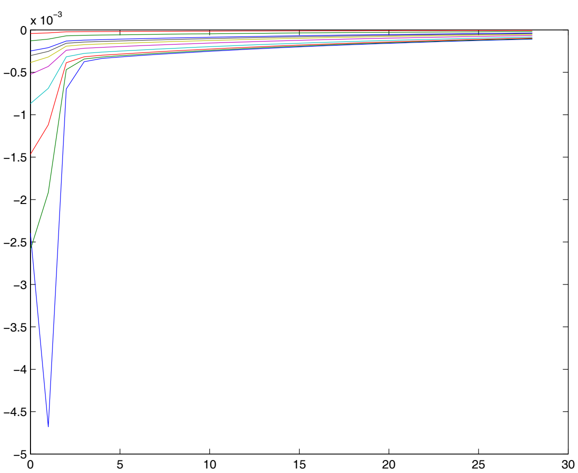

Figure 2.7.2: Impulse response of yields y_{nt} to innovation in stochastic discount factor. Bigger responses are for shorter maturity yields.

not completely flat and so reveals that the log stochastic discount factor is serially

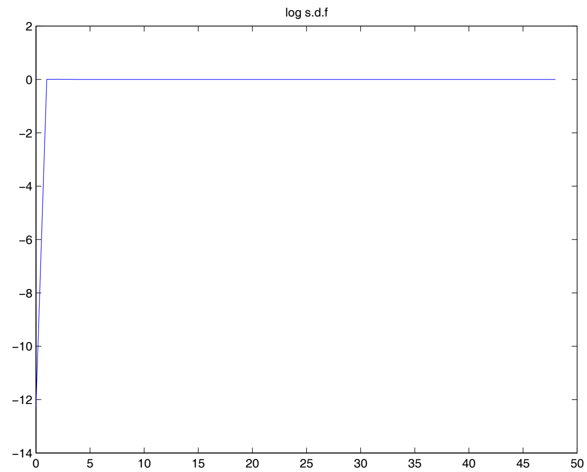

Figure 2.7.3: Impulse response of log of stochastic discount factor.

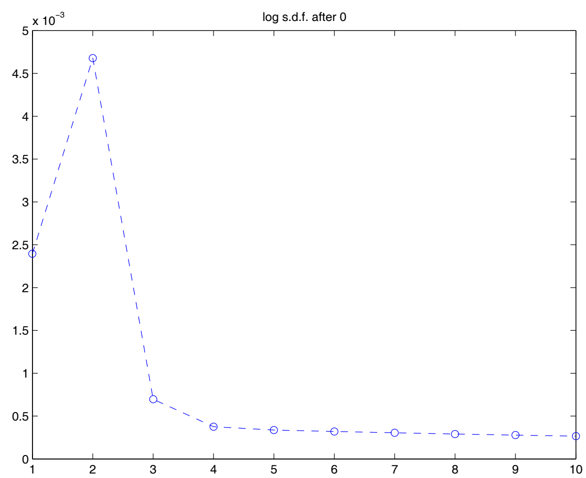

Figure 2.7.4: Impulse response of log stochastic discount factor from lag 1 on.

correlated. (Remember that the spectrum for a serially uncorrelated process – a 'white noise' – is perfectly flat.) That the spectrum is generally rising as frequency

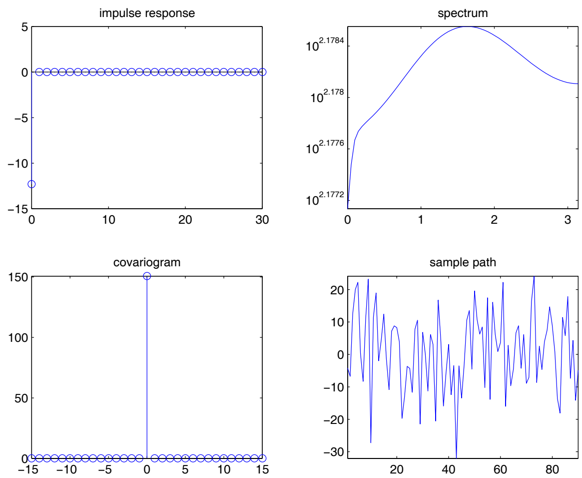

Figure 2.7.5: `bigshow2` for Backus and Zin's log stochastic discount factor.

increases from $\omega = 0$ to $\omega = \pi$ indicates that the log stochastic discount factor is *negatively* serially correlated. But the negative serial correlation is subtle so that the realization plotted in the panel on the lower right is difficult to distinguish from a white noise.

2.8. Estimation

We have shown how to map the matrices A_o, C into all of the second moments of the stationary distribution of the stochastic process $\{x_t\}$. Linear economic models typically give A_o, C as functions of a set of deeper parameters θ. We shall give examples of some such models in chapters 4 and 5. Those theories and the formulas of this chapter give us a mapping from θ to these theoretical moments of the $\{x_t\}$ process. That mapping is an important ingredient of econometric methods designed to estimate a wide class of linear rational expectations models (see Hansen and Sargent, 1980, 1981). Briefly, these methods use the following procedures for matching observations with theory. To simplify, we shall assume that in any period t that an observation is available, observations are available on the entire state x_t. As discussed in the following paragraphs, the details are more complicated if only a subset or a noisy signal of the state is observed, though the basic principles remain the same.

Given a sample of observations for $\{x_t\}_{t=0}^T \equiv x_t, t = 0, \ldots, T$, the likelihood function is defined as the joint probability distribution $f(x_T, x_{T-1}, \ldots, x_0)$. The likelihood function can be *factored* using

$$\begin{aligned} f(x_T, \ldots, x_0) = f(x_T | x_{T-1}, \ldots, x_0) \, f(x_{T-1} | x_{T-2}, \ldots, x_0) \cdots \\ f(x_1 | x_0) \, f(x_0), \end{aligned} \tag{2.8.1}$$

where in each case f denotes an appropriate probability distribution. For system (2.4.1), S$f(x_{t+1}|x_t, \ldots, x_0) = f(x_{t+1}|x_t)$, which follows from the Markov property possessed by equation (2.4.1). Then the likelihood function has the recursive form

$$f(x_T, \ldots, x_0) = f(x_T | x_{T-1}) \, f(x_{T-1} | x_{T-2}) \cdots f(x_1 | x_0) \, f(x_0). \tag{2.8.2}$$

If we assume that the w_t's are Gaussian, then the conditional distribution $f(x_{t+1}|x_t)$ is Gaussian with mean $A_o x_t$ and covariance matrix CC'. Thus, under the Gaussian distribution, the log of the conditional density of x_{t+1} becomes

$$\begin{aligned} \log f(x_{t+1} | x_t) = -.5 \log(2\pi) - .5 \det \ (CC') \\ - .5 \left(x_{t+1} - A_o x_t\right)' \left(CC'\right)^{-1} \left(x_{t+1} - A_o x_t\right) \end{aligned} \tag{2.8.3}$$

Given an assumption about the distribution of the initial condition x_0, equations (2.8.2) and (2.8.3) can be used to form the likelihood function of a sample of observations on $\{x_t\}_{t=0}^T$. One computes maximum likelihood estimates by using a hill-climbing algorithm to maximize the likelihood function with respect to the free parameters A_o, C.

When observations of only a subset of the components of x_t are available, we need to go beyond the likelihood function for $\{x_t\}$. One approach uses filtering methods to build up the likelihood function for the subset of observed variables.[25] We describe the Kalman filter in chapter 5 and the appendix on filtering and control, chapter 5.[26]

25 See Hamilton (1994) or Hansen and Sargent (in press).

26 See Hansen (1982), Eichenbaum (1991), Christiano and Eichenbaum (1992), Burnside, Eichenbaum, and Rebelo (1993), and Burnside and Eichenbaum (1996a, 1996b) for alternative estimation strategies.

2.9. Concluding remarks

In addition to giving us tools for thinking about time series, the Markov chain and the stochastic linear difference equation have each introduced us to the notion of the state vector as a description of the present position of a system.[27] Subsequent chapters use both Markov chains and stochastic linear difference equations. In the next chapter we study decision problems in which the goal is optimally to manage the evolution of a state vector that can be partially controlled.

Exercises

Exercise 2.1 Consider the Markov chain $(P, \pi_0) = \left(\begin{bmatrix} .9 & .1 \\ .3 & .7 \end{bmatrix}, \begin{bmatrix} .5 \\ .5 \end{bmatrix} \right)$, and a random variable $y_t = \overline{y} x_t$ where $\overline{y} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$. Compute the likelihood of the following three histories for y_t for $t = 0, 1, \ldots, 4$:

a. $1, 5, 1, 5, 1$.

b. $1, 1, 1, 1, 1$.

c. $5, 5, 5, 5, 5$.

Exercise 2.2 Consider a two-state Markov chain. Consider a random variable $y_t = \overline{y} x_t$ where $\overline{y} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$. It is known that $E(y_{t+1} | x_t) = \begin{bmatrix} 1.8 \\ 3.4 \end{bmatrix}$ and that $E(y_{t+1}^2 | x_t) = \begin{bmatrix} 5.8 \\ 15.4 \end{bmatrix}$. Find a transition matrix consistent with these conditional expectations. Is this transition matrix unique (i.e., can you find another one that is consistent with these conditional expectations)?

Exercise 2.3 Consumption is governed by an n state Markov chain P, π_0 where P is a stochastic matrix and π_0 is an initial probability distribution. Consumption takes one of the values in the $n \times 1$ vector $\overline{c}$. A consumer ranks stochastic processes

[27] See Quah (1990) and Blundell and Preston (1998) for applications of some of the tools of this chapter and of chapter 5 to studying some puzzles associated with a permanent income model.

of consumption $t = 0, 1 \ldots$ according to

$$E \sum_{t=0}^{\infty} \beta^t u(c_t)$$

where E is the mathematical expectation and $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$ for some parameter $\gamma \geq 1$. Let $u_i = u(\overline{c}_i)$. Let $v_i = E[\sum_{t=0}^{\infty} \beta^t u(c_t) | x_0 = \overline{e}_i]$ and $V = Ev$, where $\beta \in (0,1)$ is a discount factor.

a. Let u and v be the $n \times 1$ vectors whose ith components are u_i and v_i, respectively. Verify the following formulas for v and V: $v = (I - \beta P)^{-1} u$, and $V = \sum_i \pi_{0,i} v_i$.

b. Consider the following two Markov processes:
Process 1: $\pi_0 = \begin{bmatrix} .5 \\ .5 \end{bmatrix}$, $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
Process 2: $\pi_0 = \begin{bmatrix} .5 \\ .5 \end{bmatrix}$, $P = \begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix}$.
For both Markov processes, $\overline{c} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$.

Assume that $\gamma = 2.5, \beta = .95$. Compute unconditional discounted expected utility V for each of these processes. Which of the two processes does the consumer prefer? Redo the calculations for $\gamma = 4$. Now which process does the consumer prefer?

c. An econometrician observes a sample of 10 observations of consumption rates for our consumer. He knows that one of the two preceding Markov processes generates the data, but not which one. He assigns equal "prior probability" to the two chains. Suppose that the 10 successive observations on consumption are as follows: $1, 1, 1, 1, 1, 1, 1, 1, 1, 1$. Compute the likelihood of this sample under process 1 and under process 2. Denote the likelihood function
Prob(data|Model$_i$), $i = 1, 2$.

d. Suppose that the econometrician uses Bayes' law to revise his initial probability estimates for the two models, where in this context Bayes' law states:

$$\text{Prob}\,(M_i)\,|\text{data} = \frac{(\text{Prob}\,(\text{data})|M_i) \cdot \text{Prob}\,(M_i)}{\sum_j \text{Prob}\,(\text{data})|M_j \cdot \text{Prob}\,(M_j)}$$

where M_i denotes 'model i. The denominator of this expression is the unconditional probability of the data. After observing the data sample, what probabilities does the econometrician place on the two possible models?

e. Repeat the calculation in part d, but now assume that the data sample is $1, 5, 5, 1, 5, 5, 1, 5, 1, 5$.

Exercise 2.4 Consider the univariate stochastic process

$$y_{t+1} = \alpha + \sum_{j=1}^{4} \rho_j y_{t+1-j} + c w_{t+1}$$

where w_{t+1} is a scalar martingale difference sequence adapted to $J_t = [w_t, \ldots, w_1, y_0, y_{-1}, y_{-2}, y_{-3}]$, $\alpha = \mu(1 - \sum_j \rho_j)$ and the ρ_j 's are such that the matrix

$$A = \begin{bmatrix} \rho_1 & \rho_2 & \rho_3 & \rho_4 & \alpha \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

has all of its eigenvalues in modulus bounded below unity.

a. Show how to map this process into a first-order linear stochastic difference equation.

b. For each of the following examples, if possible, assume that the initial conditions are such that y_t is covariance stationary. For each case, state the appropriate initial conditions. Then compute the covariance stationary mean and variance of y_t assuming the following parameter sets of parameter values:

i. $\rho = [\,1.2 \quad -.3 \quad 0 \quad 0\,]$, $\mu = 10, c = 1$.

ii. $\rho = [\,1.2 \quad -.3 \quad 0 \quad 0\,]$, $\mu = 10, c = 2$.

iii. $\rho = [\,.9 \quad 0 \quad 0 \quad 0\,]$, $\mu = 5, c = 1$.

iv. $\rho = [\,.2 \quad 0 \quad 0 \quad .5\,]$, $\mu = 5, c = 1$.

v. $\rho = [\,.8 \quad .3 \quad 0 \quad 0\,]$, $\mu = 5, c = 1$.

Hint 1: The Matlab program `doublej.m`, in particular, the command `X=doublej(A,C*C')` computes the solution of the matrix equation $A'XA + C'C = X$. This program can be downloaded from
<ftp://zia.stanford.edu/pub/~sargent/webdocs/matlab>.

Hint 2: The mean vector is the eigenvector of A associated with a unit eigenvalue, scaled so that the mean of unity in the state vector is unity.

c. For each case in part b, compute the h_j's in $E_t y_{t+5} = \gamma_0 + \sum_{j=0}^{3} h_j y_{t-j}$.

d. For each case in part b, compute the $\tilde{h}_j$'s in $E_t \sum_{k=0}^{\infty} .95^k y_{t+k} = \sum_{j=0}^{3} \tilde{h}_j y_{t-j}$.

d. For each case in part b, compute the autocovariance $E(y_t - \mu_y)(y_{t-k} - \mu_y)$ for the three values $k = 1, 5, 10$.

Exercise 2.5 A consumer's rate of consumption follows the stochastic process

$$
\begin{aligned}
c_{t+1} &= \alpha_c + \sum_{j=1}^{2} \rho_j c_{t-j+1} + \sum_{j=1}^{2} \delta_j z_{t+1-j} + \psi_1 w_{1,t+1} \\
z_{t+1} &= \sum_{j=1}^{2} \gamma_j c_{t-j+1} + \sum_{j=1}^{2} \phi_j z_{t-j+1} + \psi_2 w_{2,t+1}
\end{aligned} \tag{1}
$$

where w_{t+1} is a 2×1 martingale difference sequence, adapted to $J_t = [\, w_t \quad \ldots w_1 \quad c_0 \quad c_{-1} \quad z_0 \quad z_{-1} \,]$, with contemporaneous covariance matrix $E w_{t+1} w_{t+1}' | J_t = I$, and the coefficients $\rho_j, \delta_j, \gamma_j, \phi_j$ are such that the matrix

$$
A = \begin{bmatrix}
\rho_1 & \rho_2 & \delta_1 & \delta_2 & \alpha_c \\
1 & 0 & 0 & 0 & 0 \\
\gamma_1 & \gamma_2 & \phi_1 & \phi_2 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

has eigenvalues bounded strictly below unity in modulus.

The consumer evaluates consumption streams according to

$$
V_0 = E_0 \sum_{t=0}^{\infty} .95^t u\left(c_t\right), \tag{2}
$$

where the one-period utility function is

$$
u\left(c_t\right) = -.5 \left(c_t - 60\right)^2. \tag{3}
$$

a. Find a formula for V_0 in terms of the parameters of the one-period utility function (3) and the stochastic process for consumption.

b. Compute V_0 for the following two sets of parameter values:

i. $\rho = [\, .8 \quad -.3 \,], \alpha_c = 1, \delta = [\, .2 \quad 0 \,], \gamma = [\, 0 \quad 0 \,], \phi = [\, .7 \quad -.2 \,],\ \psi_1 = \psi_2 = 1$.

ii. Same as for part i except now $\psi_1 = 2, \psi_2 = 1$.

Hint: Remember `doublej.m`.

Exercise 2.6 Consider the stochastic process $\{c_t, z_t\}$ defined by equations (1) in exercise 1.5. Assume the parameter values described in part b, item i. If possible, assume the initial conditions are such that $\{c_t, z_t\}$ is covariance stationary.

a. Compute the initial mean and covariance matrix that make the process covariance stationary.

b. For the initial conditions in part a, compute numerical values of the following population linear regression:

$$c_{t+2} = \alpha_0 + \alpha_1 z_t + \alpha_2 z_{t-4} + w_t$$

where $Ew_t \begin{bmatrix} 1 & z_t & z_{t-4} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$.

Exercise 2.7 Get the Matlab programs `bigshow2.m` and `freq.m` from <ftp://zia.stanford.edu/pub/~sargent/webdocs/matlab>. Use `bigshow2` to compute and display a simulation of length 80, an impulse response function, and a spectrum for each of the following scalar stochastic processes y_t. In each of the following, w_t is a scalar martingale difference sequence adapted to its own history and the initial values of lagged y's.

a. $y_t = w_t$.

b. $y_t = (1 + .5L)w_t$.

c. $y_t = (1 + .5L + .4L^2)w_t$.

d. $(1 - .999L)y_t = (1 - .4L)w_t$.

e. $(1 - .8L)y_t = (1 + .5L + .4L^2)w_t$.

f. $(1 + .8L)y_t = w_t$.

g. $y_t = (1 - .6L)w_t$.

Study the output and look for patterns. When you are done, you will be well on your way to knowing how to read spectral densities.

Exercise 2.8 This exercise deals with Cagan's money demand under rational expectations. A version of Cagan's (1956) demand function for money is

$$m_t - p_t = -\alpha \left(p_{t+1} - p_t\right), \alpha > 0, \ t \geq 0, \tag{1}$$

where m_t is the log of the nominal money supply and p_t is the price level at t. Equation (1) states that the demand for real balances varies inversely with the expected rate of inflation, $(p_{t+1} - p_t)$. There is no uncertainty, so the expected inflation rate equals the actual one. The money supply obeys the difference equation

$$(1 - L)(1 - \rho L) m_t^s = 0 \tag{2}$$

subject to initial condition for m_{-1}^s, m_{-2}^s. In equilibrium,

$$m_t \equiv m_t^s \quad \forall t \geq 0 \tag{3}$$

(i.e., the demand for money equals the supply). For now assume that

$$|\rho\alpha / (1 + \alpha)| < 1. \tag{4}$$

An *equilibrium* is a $\{p_t\}_{t=0}^{\infty}$ that satisfies equations (1), (2), and (3) for all t.

a. Find an expression an equilibrium p_t of the form

$$p_t = \sum_{j=0}^{n} w_j m_{t-j} + f_t. \tag{5}$$

Please tell how to get formulas for the w_j for all j and the f_t for all t.

b. How many equilibria are there?

c. Is there an equilibrium with $f_t = 0$ for all t?

d. Briefly tell where, if anywhere, condition (4) plays a role in your answer to part a.

e. For the parameter values $\alpha = 1, \rho = 1$, compute and display all the equilibria.

Exercise 2.9 The $n \times 1$ state vector of an economy is governed by the linear stochastic difference equation

$$x_{t+1} = A x_t + C_t w_{t+1} \tag{1}$$

where C_t is a possibly time varying matrix (known at t) and w_{t+1} is an $m \times 1$ martingale difference sequence adapted to its own history with $E w_{t+1} w_{t+1}' | J_t = I$, where $J_t = [\, w_t \quad \ldots \quad w_1 \quad x_0 \,]$. A scalar one-period payoff p_{t+1} is given by

$$p_{t+1} = P x_{t+1} \tag{2}$$

The stochastic discount factor for this economy is a scalar m_{t+1} that obeys

$$m_{t+1} = \frac{Mx_{t+1}}{Mx_t}. \tag{3}$$

Finally, the price at time t of the one-period payoff is given by $q_t = f_t(x_t)$, where f_t is some possibly time-varying function of the state. That m_{t+1} is a stochastic discount factor means that

$$E\left(m_{t+1}p_{t+1}|J_t\right) = q_t. \tag{4}$$

a. Compute $f_t(x_t)$, describing in detail how it depends on A and C_t.

b. Suppose that an econometrician has a time series data set $X_t = [\, z_t \quad m_{t+1} \quad p_{t+1} \quad q_t \,]$, for $t = 1, \ldots, T$, where z_t is a strict subset of the variables in the state x_t. Assume that investors in the economy see x_t even though the econometrician only sees a subset z_t of x_t. Briefly describe a way to use these data to test implication (4). (Possibly but perhaps not useful hint: recall the law of iterated expectations.)

Exercise 2.10 Let P be a transition matrix for a Markov chain. Suppose that P' has two distinct eigenvectors π_1, π_2 corresponding to unit eigenvalues of P'. Prove for any $\alpha \in [0,1]$ that $\alpha\pi_1 + (1-\alpha)\pi_2$ is an invariant distribution of P.

Exercise 2.11 Consider a Markov chain with transition matrix

$$P = \begin{bmatrix} 1 & 0 & 0 \\ .2 & .5 & .3 \\ 0 & 0 & 1 \end{bmatrix}$$

with initial distribution $\pi_0 = [\, \pi_{1,0} \quad \pi_{2,0} \quad \pi_{3,0} \,]'$. Let $\pi_t = [\, \pi_{1t} \quad \pi_{2t} \quad \pi_{3t} \,]'$ be the distribution over states at time t. Prove that for $t > 0$

$$\begin{aligned} \pi_{1t} &= \pi_{1,0} + .2\left(\frac{1-.5^t}{1-.5}\right)\pi_{2,0} \\ \pi_{2t} &= .5^t \pi_{2,0} \\ \pi_{3t} &= \pi_{3,0} + .3\left(\frac{1-.5^t}{1-.5}\right)\pi_{2,0}. \end{aligned}$$

Exercise 2.12 Let P be a transition matrix for a Markov chain. For $t = 1, 2, \ldots$, prove that the jth column of P^t is the distribution across states at t when the initial distribution is $\pi_{j,0} = 1, \pi_{i,0} = 0 \forall i \neq j$.

Exercise 2.13 A household has preferences over consumption processes $\{c_t\}_{t=0}^{\infty}$ that are ordered by

$$-.5 \sum_{t=0}^{\infty} \beta^t \left[(c_t - 30)^2 + .000001 b_t^2 \right] \tag{2.1}$$

where $\beta = .95$. The household chooses a consumption, borrowing plan to maximize (2.1) subject to the sequence of budget constraints

$$c_t + b_t = \beta b_{t+1} + y_t \tag{2.2}$$

for $t \geq 0$, where b_0 is an initial condition, where β^{-1} is the one period gross risk-free interest rate, b_t is the household's one-period debt that is due in period t, and y_t is its labor income, which obeys the second order autoregressive process

$$\left(1 - \rho_1 L - \rho_2 L^2\right) y_{t+1} = (1 - \rho_1 - \rho_2)\, 5 + .05 w_{t+1} \tag{2.3}$$

where $\rho_1 = 1.3, \rho_2 = -.4$.

a. Define the *state* of the household at t as $x_t = [\,1 \quad b_t \quad y_t \quad y_{t-1}\,]'$ and the *control* as $u_t = (c_t - 30)$. Then express the transition law facing the household in the form (2.4.22). Compute the eigenvalues of A. Compute the zeros of the characteristic polynomial $(1 - \rho_1 z - \rho_2 z^2)$ and compare them with the eigenvalues of A. (**Hint:** To compute the zeros in Matlab, set $a = [\,.4 \quad -1.3 \quad 1\,]$ and call `roots(a)`. The zeros of $(1 - \rho_1 z - \rho_2 z^2)$ equal the *reciprocals* of the eigenvalues of the associated A.)

b. Write a Matlab program that uses the Howard improvement algorithm (2.4.30) to compute the household's optimal decision rule for $u_t = c_t - 30$. Tell how many iterations it takes for this to converge (also tell your convergence criterion).

c. Use the household's optimal decision rule to compute the law of motion for x_t under the optimal decision rule in the form

$$x_{t+1} = (A - BF^*)\, x_t + C w_{t+1},$$

where $u_t = -F^* x_t$ is the optimal decision rule. Using Matlab, compute the impulse response function of $[\,c_t \quad b_t\,]'$ to w_{t+1}. Compare these with the theoretical expressions (2.6.18).

Exercise 2.14 Consider a Markov chain with transition matrix

$$P = \begin{bmatrix} .5 & .5 & 0 & 0 \\ .1 & .9 & 0 & 0 \\ 0 & 0 & .9 & .1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

with state space $X = \{e_i, i = 1, \ldots, 4\}$ where e_i is the ith unit vector. A random variable y_t is a function $y_t = [\,1 \quad 2 \quad 3 \quad 4\,]\, x_t$ of the underlying state.

a. Find all stationary distributions of the Markov chain.

b. Is the Markov chain ergodic?

c. Compute all possible limiting values of the sample mean $\frac{1}{T}\sum_{t=0}^{T-1} y_t$ as $T \to \infty$.

Exercise 2.15 Suppose that a scalar is related to a scalar white noise w_t with variance 1 by $y_t = h(L)w_t$ where $h(L) = \sum_{j=0}^{\infty} L^j h_j$ and $\sum_{j=0}^{\infty} h_j^2 < +\infty$. Then a special case of formula (2.5.6) coupled with the observer equation $y_t = Gx_t$ implies that the spectrum of y is given by

$$S_y(\omega) = h(\exp(-i\omega))\, h(\exp(i\omega)) = |h(\exp(-i\omega))|^2$$

where $h(\exp(-i\omega)) = \sum_{j=0}^{\infty} h_j \exp(-i\omega j)$.

In a famous paper, Slutsky investigated the consequences of applying the following filter to white noise: $h(L) = (1+L)^n(1-L)^m$ (i.e., the convolution of n two period moving averages with m difference operators). Compute and plot the spectrum of y for $\omega \in [-\pi, \pi]$ for the following choices of m, n:

a. $m = 10, n = 10$.

b. $m = 10, n = 40$.

c. $m = 40, n = 10$.

d. $m = 120, n = 30$.

e. Comment on these results.

Hint: Notice that $h(\exp(-i\omega)) = (1 + \exp(-i\omega))^n(1 - \exp(-i\omega))^m$.

Exercise 2.16 Consider an n-state Markov chain with state space $X = \{e_i, i = 1, \ldots, n\}$ where e_i is the ith unit vector. Consider the indicator variable $I_{it} = e_i x_t$ which equals one if $x_t = e_i$ and 0 otherwise. Suppose that the chain has a unique stationary distribution and that it is ergodic. Let π be the stationary distribution.

a. Verify that $EI_{it} = \pi_i$.

b. Prove that

$$\frac{1}{T}\sum_{t=0}^{T-1} I_{it} = \pi_i$$

as $T \to \infty$ with probability one with respect to the stationary distribution π.

Exercise 2.17 **(Lake model)**

A worker can be in one of two states, state 1 (unemployed) or state 2 (employed). At the beginning of each period, a previously unemployed worker has probability $\lambda = \int_{\bar{w}}^{B} dF(w)$ of becoming employed. Here $\bar{w}$ is his reservation wage and $F(w)$ is the c.d.f. of a wage offer distribution. We assume that $F(0) = 0, F(B) = 1$. At the beginning of each period an unemployed worker draws one and only one wage offer from F. Successive draws from F are i.i.d. The worker's decision rule is to accept the job if $w \geq \bar{w}$, and otherwise to reject it and remain unemployed one more period. Assume that $\overline{w}$ is such that $\lambda \in (0,1)$. At the beginning of each period, a previously employed worker is fired with probability $\delta \in (0,1)$. Newly fired workers must remain unemployed for one period before drawing a new wage offer.

a. Let the state space be $X = \{e_i, i = 1,2\}$ where e_i is the ith unit vector. Describe the Markov chain on X that is induced by the description above. Compute all stationary distributions of the chain. Is the chain ergodic?

b. Suppose that $\lambda = .05, \delta = .25$. Compute a stationary distribution. Compute the fraction of his life that an infinitely lived worker would spend unemployed.

c. Drawing the initial state from the stationary distribution, compute the joint distribution $g_{ij} = \text{Prob}(x_t = e_i, x_{t-1} = e_j)$ for $i = 1,2, j = 1,2$.

d. Define an indicator function by letting $I_{ij,t} = 1$ if $x_t = e_i, x_{t-1} = e_j$ at time t, and 0 otherwise. Compute

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} I_{ij,t}$$

for all four i,j combinations.

e. Building on your results in part d, construct method of moment estimators of λ and δ. Assuming that you know the wage offer distribution F, construct a method of moments estimator of the reservation wage $\bar{w}$.

f. Compute maximum likelihood estimators of λ and δ.

g. Compare the estimators you derived in parts e and f.

h. *Extra credit.* Compute the asymptotic covariance matrix of the maximum likelihood estimators of λ and δ.

Exercise 2.18 (**random walk**)

A Markov chain has state space $X = \{e_i, i = 1, \ldots, 4\}$ where e_i is the unit vector and transition matrix

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ .5 & 0 & .5 & 0 \\ 0 & .5 & 0 & .5 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

A random variable $y_t = \overline{y} x_t$ is defined by $\overline{y} = [\,1 \quad 2 \quad 3 \quad 4\,]$.

a. Find all stationary distributions of this Markov chain.

b. Is this chain ergodic? Compute invariant functions of P.

c. Compute $E[y_{t+1}|x_t]$ for $x_t = e_i, i = 1, \ldots, 4$.

d. Compare your answer to part (c) with (2.2.9). Is $y_t = \overline{y}' x_t$ invariant? If not, what hypothesis of Theorem 2.2.2 is violated?

d. The stochastic process $y_t = \overline{y}' x_t$ is evidently a bounded martingale. Verify that y_t converges almost surely to a constant. To what constant(s) does it converge?

A. A linear difference equation

This appendix describes the solution of a linear first-order scalar difference equation. First, let $|\lambda| < 1$, and let $\{u_t\}_{t=-\infty}^{\infty}$ be a bounded sequence of scalar real numbers. Then

$$(1 - \lambda L)\, y_t = u_t, \forall t \tag{2.A.1}$$

has the solution

$$y_t = (1 - \lambda L)^{-1} u_t + k\lambda^t \tag{2.A.2}$$

for any real number k. You can verify this fact by applying $(1 - \lambda L)$ to both sides of equation $(2.A.2)$ and noting that $(1 - \lambda L)\lambda^t = 0$. To pin down k we need one condition imposed from outside (e.g., an initial or terminal condition) on the path of y.

Now let $|\lambda| > 1$. Rewrite equation $(2.A.1)$ as

$$y_{t-1} = \lambda^{-1} y_t - \lambda^{-1} u_t, \forall t \tag{2.A.3}$$

or

$$\left(1-\lambda^{-1}L^{-1}\right)y_t=-\lambda^{-1}u_{t+1}. \tag{2.A.4}$$

A solution is

$$y_t=-\lambda^{-1}\left(\frac{1}{1-\lambda^{-1}L^{-1}}\right)u_{t+1}+k\lambda^t \tag{2.A.5}$$

for any k. To verify that this is a solution, check the consequences of operating on both sides of equation $(2.A.5)$ by $(1-\lambda L)$ and compare to $(2.A.1)$.

Solution $(2.A.2)$ exists for $|\lambda|<1$ because the distributed lag in u converges. Solution $(2.A.5)$ exists when $|\lambda|>1$ because the distributed lead in u converges. When $|\lambda|>1$, the distributed lag in u in $(2.A.2)$ may diverge, so that a solution of this form does not exist. The distributed lead in u in $(2.A.5)$ need not converge when $|\lambda|<1$.

Chapter 3.
Dynamic Programming

This chapter introduces basic ideas and methods of dynamic programming.[1] It sets out the basic elements of a recursive optimization problem, describes the functional equation (the Bellman equation), presents three methods for solving the Bellman equation, and gives the Benveniste-Scheinkman formula for the derivative of the optimal value function. Let's dive in.

3.1. Sequential problems

Let $\beta \in (0,1)$ be a discount factor. We want to choose an infinite sequence of "controls" $\{u_t\}_{t=0}^{\infty}$ to maximize

$$\sum_{t=0}^{\infty} \beta^t r\left(x_t, u_t\right), \tag{3.1.1}$$

subject to $x_{t+1} = g(x_t, u_t)$, with x_0 given. We assume that $r(x_t, u_t)$ is a concave function and that the set $\{(x_{t+1}, x_t) : x_{t+1} \leq g(x_t, u_t), u_t \in R^k\}$ is convex and compact. Dynamic programming seeks a time-invariant *policy function* h mapping the *state* x_t into the control u_t, such that the sequence $\{u_s\}_{s=0}^{\infty}$ generated by iterating the two functions

$$\begin{aligned} u_t &= h\left(x_t\right) \\ x_{t+1} &= g\left(x_t, u_t\right), \end{aligned} \tag{3.1.2}$$

starting from initial condition x_0 at $t = 0$ solves the original problem. A solution in the form of equations (3.1.2) is said to be *recursive*. To find the policy function h we need to know another function $V(x)$ that expresses the optimal value of the original problem, starting from an arbitrary initial condition $x \in X$. This is called the *value*

[1] This chapter is written in the hope of getting the reader to start using the methods quickly. We hope to promote demand for further and more rigorous study of the subject. In particular see Bertsekas (1976), Bertsekas and Shreve (1978), Stokey and Lucas (with Prescott) (1989), Bellman (1957), and Chow (1981). This chapter covers much of the same material as Sargent (1987b, chapter 1).