In order to explain these principles, we will first describe a method for computing the process Z, a topic of importance in its own right. The computations will be done with dynamic programming, an algorithmic method of wide applicability. The idea in the case of American options is to work backwards in time, computing as you go the process Z. Note that (4.8) is clearly true for $t = T$, that is, $Z_T = Y_T$. Moreover, (4.9) holds for $t = T$.

Now the first iteration is to compute Z_{T-1} by taking

$$Z_{T-1} = \max\{Y_{T-1}, E_Q[Z_T B_{T-1}/B_T | \mathscr{F}_{T-1}]\} \tag{4.10}$$

noting for future reference that it will be optimal to take $\tau = T - 1$ if and only if $Z_{T-1} = Y_{T-1}$. Now (4.7) can be rewritten for $t = T - 1$ as

$$Z_{T-1} = \max\{Y_{T-1}, E_Q[Y_T B_{T-1}/B_T | \mathscr{F}_{T-1}]\}$$

so comparing this with (4.10) we realize that Z_{T-1} given by (4.10) satisfies (4.7) for $t = T - 1$. Moreover, the stopping time that is spelled out in (4.9) for $T - 1$ is the one which maximizes the right hand side of (4.7) for $t = T - 1$.

Now suppose you have computed Z_t and you know that this satisfies (4.7) and that (4.9) for time t gives a stopping time which maximizes the right hand side of (4.7) for time t. Then Z_{t-1}, as given by

$$Z_{t-1} = \max\{Y_{t-1}, E_Q[Z_t B_{t-1}/B_t | \mathscr{F}_{t-1}]\} \tag{4.11}$$

will satisfy (4.7) for $t - 1$. To see this, note that (4.7) holds for time t, so with Z_{t-1} as has been computed we have

$$\begin{aligned} Z_{t-1} &= \max\{Y_{t-1}, E_Q[\max_{\tau \in \zeta(t,T)} \{E_Q[Y_\tau B_t/B_\tau | \mathscr{F}_t] B_{t-1}/B_t | \mathscr{F}_{t-1}]\} \\ &\geqslant \max\{Y_{t-1}, E_Q[E_Q[Y_\tau B_t/B_\tau | \mathscr{F}_t] B_{t-1}/B_t | \mathscr{F}_{t-1}]\} \\ &= \max\{Y_{t-1}, E_Q[Y_\tau B_{t-1}/B_\tau | \mathscr{F}_{t-1}]\} \end{aligned} \tag{4.12}$$

for all $\tau \in \zeta(t, T)$. Hence Z_{t-1} as has been computed is greater than or equal to the right hand side of (4.7) for time $t - 1$. On the other hand, taking the stopping time in (4.9) for time $t - 1$ we see that (4.12) becomes an equality. It follows that Z_{t-1} as has been computed satisfies (4.7) for time t and that the stopping time given by (4.9) for time $t - 1$ is the one which maximizes the right hand side of (4.7) for time $t - 1$.

Hence by a mathematical induction argument we have described an algorithm for computing a process Z which satisfies (4.7), and we have verified that (4.9) gives the stopping times which achieve the maximum on the right hand side of (4.7). This algorithm is based on dynamic programming, which is simply the idea that if the optimal exercise strategy starting at time $t - 1$ calls for continuing at least one more period, then next period you will use the optimal exercise strategy starting at time t.

Example 4.1 (continued) Starting with $Z_2 = Y_2$ we compute for states ω_1 and ω_2

$$E_Q[Z_2|\mathscr{F}_1] = E_Q[Y_2|\mathscr{F}_1] = \frac{2}{3}(4) + \frac{1}{3}(1) = 3$$

Hence $Z_1 = \max\{Y_1, E_Q[Z_2|\mathscr{F}_1]\} = \max\{4, 3\} = 4$ in these same states. Similarly, in states ω_3 and ω_4 we have $E_Q[Z_2|\mathscr{F}_1] = 1/3$ and $Z_1 = \max\{0, 1/3\} = 1/3$.

For the next dynamic programming iteration, that is, for time 0, we have

$$E_Q[Z_1|\mathscr{F}_0] = E_Q Z_1 = \frac{1}{4}(4) + \frac{3}{4}(1/3) = 1.25$$

in which case $Z_0 = \max\{Y_0, E_Q Z_1\} = \max\{1, 1.25\} = 1.25$. Note Z is displayed in figure 4.3.

Looking at (4.11) it is easy to see that $Z \geqslant Y$ and that $Z_t \geqslant E_Q[Z_{t+1} B_t/B_{t+1} | \mathscr{F}_t]$ for $t = 0, 1, \ldots, T-1$. It follows from this inequality that Z/B is a Q-supermartingale. Note that the value process $Z = Z/B$ (recall $B = 1$) in figure 4.3 for example 4.1 is indeed a Q-supermartingale.

Now suppose U is another process satisfying $U \geqslant Y$ and such that U/B is a Q-supermartingale. Then

$$U_{t-1} \geqslant \max\{Y_{t-1}, E_Q[U_t B_{t-1}/B_t | \mathscr{F}_{t-1}]\}, \quad t = 1, \ldots, T \tag{4.13}$$

In particular, $U_T \geqslant Y_T = Z_T$, so taking (4.13) for $t = T$ we realize from (4.11) that $U_{t-1} \geqslant Z_{t-1}$. We can repeat this argument, working backwards in time, until we eventually obtain the conclusion that $U_0 \geqslant Z_0$. We thus conclude that Z/B is the smallest Q-supermartingale such that $Z \geqslant Y$. This completes the verification of principle (4.6).

Recall that the time t price of a European option X is given by $E_Q[XB_t/B_T|\mathscr{F}_t]$ provided X is marketable, whereas we are unable to pin down the price of X if it is not marketable. A similar situation exists for American options. If an American option Y is marketable (to be defined shortly), then its time t price is given by Z_t, as developed in (4.6). On the other hand, if Y is not marketable, then we are unable to specify its price.

The American option Y will be *marketable* or *attainable* if for each stopping time $\tau \leqslant T$ there exists a self-financing trading strategy such that the corresponding portfolio value V satisfies $V_\tau = Y_\tau$. In other words, Y will be marketable if one can *replicate* claims of the form Y_τ, and the trading strategy that corresponds to $V_\tau = Y_\tau$ is called a *replicating* or *hedging* trading strategy. This definition of marketability is stronger than necessary, since we really only need to worry about stopping times of the form (4.9). However, the approach taken here is simpler, and it is a property that can be checked without having to first compute the American option's value process Z.

Example 4.1 (continued) For the American option in figure 4.3 we will only check the stopping times of the form (4.9), that is

$$\tau(0) = \tau(1) = \begin{cases} 1, & \omega = \omega_1, \omega_2 \\ 2, & \omega = \omega_3, \omega_4 \end{cases}$$

and $\tau(2)(\omega)=2$, *all* $\omega \in \Omega$. Now $Y_{\tau(2)} = Y_2$ is the same as the European call option X that we have been studying; we know X is marketable, so there exists a trading strategy such that $V_{\tau(t)} = Y_{\tau(t)}$ holds for $t = 2$.

For $\tau \equiv \tau(0) = \tau(1)$ we have

$$Y_\tau = \begin{cases} 4, & \omega = \omega_1, \omega_2 \\ 1, & \omega = \omega_3 \\ 0, & \omega = \omega_4 \end{cases}$$

To see if there is a trading strategy such that $V_{\tau(t)} = Y_{\tau(t)}$ is satisfied for $t = 0$ and 1, we work backwards in time, trying to compute the requisite trading strategy, beginning with the largest value of τ:

$$H_0(2) + 6\,H_1(2) = Y_2(\omega_3) = 1$$
$$H_0(2) + 3\,H_1(2) = Y_2(\omega_4) = 0$$

The solution exists and is $H_0(2) = -1$ and $H_1(2) = 1/3$ in states ω_1 and ω_2. Since the trading strategy must be self-financing, this implies $V_1 = 1/3$ in the same states.

The values of $H_0(2)$ and $H_1(2)$ for states ω_3 and ω_4 do not really matter, since $\tau = 1$ in these states. All that remains is to see if there are values of $H_0(1)$ and $H_1(1)$ satisfying

$$H_0(1) + 8\,H_1(1) = Y_1(\omega_1) = Y_1(\omega_2) = 4$$
$$H_0(1) + 4\,H_1(1) = V_1(\omega_3) = V_1(\omega_4) = 1/3$$

The solution exists and is $H_0(1) = -10/3$ and $H_1(1) = 11/12$. Thus there exists a trading strategy satisfying $V_{\tau(t)} = Y_{\tau(t)}$ for t equal to 0 and 1, in which case Y is a marketable American option. Notice that the time 0 value of the portfolio we just derived is $-10/3 + 5\,(11/12) = 5/4$, which is the same as the value of Z_0 for this American option that was computed earlier.

(4.14) Suppose there exists a risk neutral probability measure Q, the process Z is as in (4.7), and the American option Y is marketable. Then Z is the value or price process for Y, and the optimal early exercise strategy $\tau(0)$ is given by (4.9).

To see that Z is the price process for Y, one can use an arbitrage argument similar to what was used for European options. Briefly, let p denote the time 0 price of Y, and suppose $p > Z_0$. Then one can sell the option for p dollars, undertake the trading strategy that replicates $Y_{\tau(0)}$ at a cost of Z_0, and invest the difference $p - Z_0$ in the bank account. Later, if the buyer exercises the option at time $t \leqslant \tau(0)$, you liquidate the portfolio, collecting Z_t dollars and paying the option buyer Y_t. Since $Z_t \geqslant Y_t$, these transactions at the exercise time will only add to your profit, thereby guaranteeing yourself a strictly positive profit overall.

On the other hand, suppose the option buyer does not exercise by time $t = \tau(0)$, where $\tau(0) < T$. Then you repeat this process, undertaking the trading strategy that replicates $Y_{\tau(t)}$ at a cost of $E_Q[Z_{t+1}B_t/B_{t+1}|\mathscr{F}_t]$, which is less than or equal to $Z_t = Y_t$ by the dynamic programming relationship (4.11). As before, if the option buyer exercises at some time $s \leqslant \tau(t)$, then the value of the portfolio will be enough to cover the payoff Y_s. If the buyer still has not exercised by time $\tau(t)$, then you repeat this process yet again, and so forth. The bottom line: you will always have enough money in your portfolio to cover the payoff, and your overall profit will be at least $p - Z_0 > 0$.

For the opposite case, suppose the time 0 price of the option satisfies $p < Z_0$. Then you buy the option for p dollars, you undertake the negative of the replicating trading strategy, thereby collecting Z_0 dollars, and you invest the difference $Z_0 - p$ in the bank account. Later you exercise the option at time $\tau(0)$, and you liquidate the replicating portfolio at the same time. Since $V_{\tau(0)} = Y_{\tau(0)}$, the amount you collect from the option seller is exactly equal to your liability on the portfolio. Meanwhile, you have $(Z_0 - p)B_{\tau(0)} > 0$ dollars in the bank account.

Hence if $p \neq Z_0$ there will exist an arbitrage opportunity, so Z_0 must be the time 0 price of the American option Y. Moreover, an optimal exercise strategy for the option buyer is given by (4.9), because any other strategy runs the risk of exercising when $Z_t(\omega) > Y_t(\omega)$, which means the option's buyer would foolishly sacrifice the amount $Z_t(\omega) - Y_t(\omega) > 0$. A similar argument will verify that Z_t is the time t price of this option and that (4.9) also gives the optimal exercise strategy starting at time t when $t > 0$.

We now turn to a new question: under what circumstances should the buyer of an American option never exercise early? Principle (4.5) gives one sufficient condition for $\tau = T$ to be the optimal exercise strategy. An adapted stochastic process $Z = \{Z_t; t = 0, 1, \ldots, T\}$ is said to be a *submartingale* (see section 3.3) if

$$E[Z_t|\mathscr{F}_s] \geqslant Z_s, \quad \textit{all}\ 0 \leqslant s \leqslant t \leqslant T,$$

that is, if $-Z$ is a supermartingale. This definition permits one to state another sufficient condition.

(4.15) If Y is a marketable American option and Y/B is a Q-submartingale, then $\tau = T$ is always an optimal exercise strategy, and the price of this American option coincides with the price of the European option $X = Y_T$.

To understand this, we need something called the *optional sampling theorem*, which says that if Y/B is a Q-submartingale, then $E_Q[Y_\tau/B_\tau] \leqslant E_Q[Y_T/B_T]$ for all stopping times $\tau \leqslant T$. Hence by (4.7) with $t = 0$ we have $Z_0 = E_Q[Y_T/B_T]$, which we recognize as the price of the European option $X = Y_T$.

Principle (4.15) provides a condition that is sometimes convenient for checking whether the American and European options' values coincide. For example, we have the following.

(4.16) In a world of non-negative interest rates and no dividends, an American call option written on an individual risky security should not be exercised early.

To see this it suffices to verify that $(S_t - e)^+/B_t = (S_t/B_t - e/B_t)^+$ is a Q-submartingale by showing for arbitrary s, $t \geqslant 0$ that

$$(S_t/B_t - e/B_t)^+ \leqslant E_Q[\,(S_{t+s}/B_{t+s} - e/B_{t+s})^+|\mathscr{F}_t]$$

To begin with, we have

$$E_Q[\,(S_{t+s}/B_{t+s} - e/B_{t+s})^+|\mathscr{F}_t] \geqslant E_Q[\,(S_{t+s}/B_{t+s} - e/B_{t+s})|\mathscr{F}_t]$$

$$= E_Q[S_{t+s}/B_{t+s}|\mathscr{F}_t] - eE_Q[1/B_{t+s}|\mathscr{F}_t] = S_t/B_t - eE_Q[1/B_{t+s}|\mathscr{F}_t]$$

where the last equality is because S/B is a Q-martingale. But $B_{t+s} \geqslant B_t$, so $1/B_t \geqslant 1/B_{t+s}$ and thus

$$E_Q[\,(S_{t+s}/B_{t+s} - e/B_{t+s})^+|\mathscr{F}_t] \geqslant S_t/B_t - eE_Q[1/B_t|\mathscr{F}_t] = S_t/B_t - e/B_t$$

Finally, since $E_Q[\,(S_{t+s}/B_{t+s} - e/B_{t+s})^+|\mathscr{F}_t] \geqslant 0$, it follows that $E_Q[\,(S_{t+s}/B_{t+s} - e/B_{t+s})^+|\mathscr{F}_t] \geqslant \max\{0, S_t/B_t - e/B_t\}$, which is exactly what we wanted to show.

It should be emphasized that a crucial part of this verification of (4.16) is the requirement that the discounted security price S/B is a Q-martingale. If the risky security pays a dividend, then the discounted, ex-dividend price of this security is not necessarily a Q-martingale, in which case the American call option based on the ex-dividend price might be worth strictly more than the corresponding European option.

Exercise 4.8 For the model in example 4.1 with $r = 0$, what is the time 0 price of an American put that has exercise price $e = 6$? Is it optimal to exercise early? If so, when? And how do you hedge this option if you sell it at time 0?

Exercise 4.9 Consider the binomial stock price model with $T = 4$, $S_0 = 20$, $u = 1.2214$, and $d = 0.8187 = u^{-1}$. The interest rate is $r = 3.82\%$. What is the time 0 price of an American put that has exercise price $e = 18$? Is it optimal to exercise early? If so, when?

Exercise 4.10 Show that if $M = \{M_t; t = 0, \ldots, T\}$ is a supermartingale, then $E[M_\tau|\mathscr{F}_t] \leqslant M_t$ for all stopping times τ with $t \leqslant \tau \leqslant T$.

Exercise 4.11 Suppose the American option $Y_t = g(S_t)$, where g is a convex function with $g(0) = 0$. Let the interest rate be constant with $r \geqslant 0$. Show that $\tau = T$ is an optimal exercise strategy. (Hint: use (4.15) together with *Jensen's inequality*, which says that $g(E[X|\mathscr{F}]) \leqslant E[g(X)|\mathscr{F}]$ for any convex function g, any random variable X, and any algebra $\mathscr{F}$.)

Exercise 4.12 Suppose b is a positive constant and the interest rates are non-negative. Show that $-b/B_t$ is a Q-submartingale. If $M_1(t), \ldots, M_J(t)$ are Q-submartingales and $m_1, \ldots, m_J$ are positive scalars, then show $m_1M_1 + \ldots + m_JM_J$ is a Q-submartingale. Hence show for American options of the form

$$Y_t = m_0S_t + m_1(S_t - e_1)^+ + \ldots + m_J(S_t - e_J)^+ - b$$

where $m_1 \geqslant 0\ldots$, $m_J \geqslant 0$, and $b \geqslant 0$ (but m_0 can be negative), that it is optimal to not exercise early.

4.4 Complete and Incomplete Markets

In order to price a contingent claim X you need to worry about whether it is marketable, that is, whether there exists a self-financing trading strategy such that $V_T = X$. In this regard, the situation is virtually the same as with single period models. This section will explain how the single period results generalize to multiperiod models and give special attention to the new issue of marketable American options. Throughout this section, of course, it will be assumed that there exists a risk neutral probability measure Q.

The model is said to be *complete* if every contingent claim is marketable; otherwise, the model is said to be *incomplete*. For single period models, we saw there were two characterizations of completeness: either (1.22), the number of independent vectors in a certain matrix, denoted A, equals the number of states, or (1.24), the risk neutral probability measure is unique. Both characterizations generalize to multiperiod models, as can be seen by decomposing the multiperiod model into a network of single period models, as was done in section 3.3. In particular, think of the information tree corresponding to the multiperiod model, with one underlying single period model at each node of this network.

If the multiperiod model is complete, then for an arbitrary contingent claim X you can always work backwards in time, as was explained in section 4.1, to compute the trading strategy that generates X. This means the matrix A for each underlying single period model must have the requisite number of independent columns, or the computational procedure may fail. Thus every underlying single period model must be complete. Conversely, if every underlying single period model is complete, then the computational procedure for the multiperiod model will always succeed, and we have the following.

(4.17) The multiperiod model is complete if and only if every underlying single period model is complete.

In particular, the multiperiod model is complete if and only if the A-matrix corresponding to every underlying single period model has the requisite number of independent columns. Now this generalization of (1.22) is a rather ugly and impractical characterization, so it will not be highlighted.

On the other hand, the generalization of (1.24) is rather nice. Every underlying single period model is complete if and only if each underlying single period model has a unique risk neutral 'conditional' probability measure. In view of the construction that was developed in section 3.3 of the risk neutral probability measure for the multiperiod model, every underlying single period model has a unique risk neutral 'conditional' probability measure if and only if the risk neutral probability measure for the multiperiod model is unique. Hence we have the following.

(4.18) The multiperiod model is complete if and only if the risk neutral probability measure Q is unique.

Example 4.1 (continued) This model is complete, because the risk neutral probability measure is unique.

Example 4.4 (continued) This model is complete, because the risk neutral probability measure is unique.

Example 4.9 (continued) The binomial model studied in section 4.2 is complete, because the risk neutral probability measure is unique.

In an incomplete model, therefore, there will exist at least two, and thus many, risk neutral probability measures. If X is an attainable contingent claim in an incomplete model, then its time 0 price will equal V_0 for the trading strategy which generates X. Since $V_0 = E_Q[X/B_T]$ for every risk neutral probability measure Q, we realize that for marketable contingent claims the quantity $E_Q[X/B_T]$ is constant over the risk neutral probability measures, that is, over all $Q \in \mathbb{M}$.

To show the converse, it suffices to suppose the contingent claim X is not attainable and then demonstrate that $E_Q[X/B_T]$ does not take the same value for all $Q \in \mathbb{M}$. The argument will only be sketched. Again, think of the network of underlying single period models, and suppose you are trying to compute the replicating trading strategy by working backwards in time. This procedure will break down at some single period model. As seen in section 1.5, this means there will exist at least two sets of risk neutral 'conditional' probabilities for this single period model giving different conditional expectations to X/B_T. It follows that there will exist at least two martingale measures Q giving different values to $E_Q[X/B_T]$.

In summary, we have the following generalization of (1.23), true for both complete and incomplete models.

(4.19) The contingent claim X is attainable if and only if $E_Q[X/B_T]$ takes the same value for every $Q \in \mathbb{M}$.

Turning to American options, fortunately there is the following.

(4.20) If the model is complete, then every American option is marketable.

To see this it suffices to take an arbitrary non-negative adapted process Y and an arbitrary stopping time τ and show that there exists a self-financing trading strategy such that $V_\tau = Y_\tau$. To do this we use a trick: we consider a trading strategy, say $\hat{H}$, which starts at time τ with the Y_τ dollars, all of which is put into and kept in the bank account until time T. Under such a strategy there will be $Y_\tau B_T/B_\tau$ dollars at time T. Meanwhile, since the model is complete, we know there exists a trading strategy, say H, which starts at time 0 and satisfies $V_T = Y_\tau B_T/B_\tau$. Since the time T values of the two portfolios are the same, we realize that the time τ values of the portfolios under H and $\hat{H}$ must coincide, that is, $V_\tau = \hat{V}_\tau = Y_\tau$. Hence H is the strategy we were seeking.

If the market is incomplete, then an arbitrary American option may or may not be marketable. In view of (4.19) and the preceding paragraph, we have the following.

(4.21) The American option Y is attainable if and only if, for each stopping time τ in (4.9), $E_Q[Y_\tau/B_\tau]$ takes the same value for all $Q \in \mathbb{M}$.

Unfortunately, this condition is not so easy to check in particular cases.

Example 4.10 Suppose $K = 5$, $N = 1$, $T = 2$, $r = 0$, the filtration $\mathscr{F}_t$ is generated by the price process S, and S is as shown:

ω	S_0	S_1	S_2	Q
ω_1	5	8	9	$q/4$
ω_2	5	8	7	$(2-3q)/4$
ω_3	5	8	6	$(2q-1)/4$
ω_4	5	4	6	$1/4$
ω_5	5	4	3	$1/2$

The martingale measure Q was computed and is displayed in this table; here q is any scalar satisfying $1/2 < q < 2/3$. This model is not complete, because the martingale measure is not unique.

The contingent claim X is marketable if and only if

$$\begin{aligned} E_q X &= [qX_1 + (2-3q)X_2 + (2q-1)X_3 + X_4 + 2X_5]/4 \\ &= q[X_1 - 3X_2 + 2X_3]/4 + X_2/2 - X_3/4 + X_4/4 + X_5/2 \end{aligned}$$

is constant with respect to q, where here X_i denotes $X(\omega_i)$. Hence the contingent claim X is marketable if and only if

$$X_1 - 3X_2 + 2X_3 = 0$$

To check whether the American option Y is marketable, we certainly need to check whether this equation is satisfied with $X = Y_2$. In addition, we need to check whether, for relevant stopping times τ, $E_q Y_\tau$ is constant with respect to q. There are only two stopping times of

interest: either $\tau_a \equiv 1_{\{S_1=8\}} + 2 \cdot 1_{\{S_1=4\}}$ or $\tau_b \equiv 2 \cdot 1_{\{S_1=8\}} + 1_{\{S_1=4\}}$. Applying (4.21), we compute

$$E_q Y_{\tau_a} = Y_1(\omega_1)/4 + Y_2(\omega_4)/4 + Y_2(\omega_5)/2$$

(recall Y is adapted, so $Y_1(\omega_1) = Y_1(\omega_2) = Y_1(\omega_3)$). This is clearly constant with respect to q. Similarly, we compute

$$\begin{aligned} E_q Y_{\tau_b} &= Y_2(\omega_1)q/4 + Y_2(\omega_2)(2-3q)/4 + Y_2(\omega_3)(2q-1)/4 + 3Y_1(\omega_4)/4 \\ &= \frac{q}{4}[Y_2(\omega_1) - 3Y_2(\omega_2) + 2Y_2(\omega_3)] + Y_2(\omega_2)/2 - Y_2(\omega_3)/4 + 3Y_1(\omega_4)/4 \end{aligned}$$

The condition that this be constant with respect to q leads to the same requirement as before: the American option Y is marketable if and only if $X_1 - 3X_2 + 2X_3 = 0$ is satisfied with $X = Y_2$, that is, if and only if the European option $X = Y_2$ is marketable.

Exercise 4.13 Consider the model in example 4.10. For what values of the exercise price e is the call option marketable? For what values of e is the put option marketable?

4.5 Forward Prices and Cash Stream Valuation

Suppose you agree at time t to acquire a security at time τ, paying for it at time τ with an amount you negotiated at time t. Or suppose you want to purchase a cash stream of future receipts in exchange for your payment at time t. As will be discussed in this section, the prices of these derivative securities can be evaluated with arbitrage pricing theory.

Actually, these derivative securities can only be evaluated if they can be replicated with trading strategies, so for ease of exposition it will be assumed throughout this section that the securities market model is complete.

First consider a cash stream which consists of ΔD_s dollars received at time s, where $t < s \leqslant T$. This is just a time s contingent claim, so its time t value is $E_Q[\Delta D_s B_t/B_s|\mathscr{F}_t]$. If ΔD_s is deterministic (or even is $\mathscr{F}_t$ measurable), then this simplifies to $\Delta D_s E_Q[B_t/B_s|\mathscr{F}_t]$. If, in addition, the interest rate $r \geqslant 0$ is constant, then this simplifies further to $\Delta D_s(1+r)^{t-s}$. All these expressions are often called the time t *present value* of ΔD_s.

Next consider a cash stream $\Delta D_{t+1}, \ldots, \Delta D_T$. For example, these receipts may be the dividends associated with one unit of a security S. In view of the preceding paragraph, the time t present value of this cash stream is simply

$$\sum_{s=t+1}^{T} E_Q[\Delta D_s B_t/B_s|\mathscr{F}_t].$$

In the special case where this cash stream comes from the dividend process D associated with the security S, it follows from (3.28) that the time t present value of this cash stream is

$$S_t - E_Q[S_\tau B_t/B_\tau|\mathscr{F}_t].$$

These present value expressions are of fundamental importance in finance, especially when the interest rate is constant.

Sometimes two parties may each have a cash stream which, although featuring different receipts at various future times, have identical time t present values. In such a case, these two parties may find it advantageous to trade their cash streams. Such a transaction is called a *swap*. Swaps are also made when the time t present values of the two cash streams are different, simply by having one party pay the other at time t an amount equal to the difference between the two time t present values.

Turning to the subject of forward prices, suppose at time t you agree to purchase one unit of security S from a broker. No money or security units are exchanged at time t, but the agreement you made at time t stipulates that at the future time $\tau, t < \tau \leqslant T$, you will receive one unit of security S and you will pay your broker O_t dollars (note: the notation 'O_t' comes from *fOrward*). The question is: is there a correct value for the *forward price* O_t?

It should not come as a surprise to learn that the answer to this question is yes. The idea is to realize that holding the forward agreement plus O_t dollars at time τ is equivalent to holding one unit of security S at the same time. To be specific, in the first strategy you enter the forward agreement at time t, at the same time you replicate the contingent claim which pays O_t dollars at time τ, and at time τ you fulfill the forward agreement by paying O_t dollars and receiving the security. The time t cost of replicating O_t is simply the present value of O_t, that is,

$$E_Q[O_t B_t/B_\tau|\mathscr{F}_t] = O_t E_Q[B_t/B_\tau|\mathscr{F}_t].$$

The second strategy is worth S_τ at time τ, and its time t present value is simply S_t (by the definition of the martingale measure Q). But the first strategy is also worth S_τ at time τ, so by the law of one price the time t values of the two strategies must be equal, that is, $S_t = O_t E_Q[B_t/B_\tau|\mathscr{F}_t]$. In other words, we have the following.

(4.22) The time t forward price O_t of security S, which is received and paid for at time $\tau > t$ and which pays no dividends, is

$$O_t = \frac{S_t}{E_Q[B_t/B_\tau|\mathscr{F}_t]}$$

It is worth pointing out that this expression for the forward price is not necessarily true if the security S pays dividends. This is because (4.22) was derived by assuming S/B is a Q-martingale, an assumption that is false when the security pays a dividend. However, principle (4.22) can easily be generalized to the case of a dividend-paying security.

To derive a generalization, we need to concoct and then price the trading strategy which replicates the time τ contingent claim equal to S_τ. Here it is:

- At time t purchase one unit of the security by paying S_t.
- At time t borrow $E_Q[\Delta D_{t+1} B_t / B_{t+1} | \mathscr{F}_t]$ dollars by undertaking the negative of the trading strategy that replicates the time $t+1$ receipt ΔD_{t+1}. Then at time $t+1$ use the dividend payment ΔD_{t+1} to settle the liability under this strategy.

⋮

- At time t borrow $E_Q[\Delta D_\tau B_t / B_\tau | \mathscr{F}_t]$ dollars by undertaking the negative of the trading strategy that replicates the time τ receipt ΔD_τ. Then at time τ use the dividend payment ΔD_τ to settle the liability under this strategy.

The time t value of this replicating portfolio must be

$$S_t - \sum_{s=t+1}^{\tau} E_Q[\Delta D_s B_t / B_s | \mathscr{F}_t]$$

Equating this with the time t present value of the time τ receipt O_t, we obtain the following conclusion.

(4.23) The time t forward price O_t of security S, which is received and paid for at time $\tau > t$, and which has the dividend process D, is

$$O_t = \frac{S_t}{E_Q[B_t / B_\tau | \mathscr{F}_t]} - \sum_{s=t+1}^{\tau} \frac{E_Q[\Delta D_s B_t / B_s | \mathscr{F}_t]}{E_Q[B_t / B_\tau | \mathscr{F}_t]}$$

Sometimes forward prices are associated with assets which cannot be classified as securities because it is impossible to sell them short. For example, most agricultural commodities, such as live hogs, bushels of corn, and bales of cotton, cannot be sold short, perhaps because there is no market to borrow these physical goods from the farmers. Consequently, the usual arbitrage arguments break down, and one cannot be sure whether the discounted cash (or *spot*) prices of these assets are martingales under any probability measure.

For many such assets it is perfectly feasible to purchase them, provided you are willing to pay a *carrying cost* of c dollars per unit per period (perhaps this is an inventory storage charge). This is like a security that pays a dividend, only the dividend process satisfies $\Delta D_t = -c$ for all t.

Let A_t denote the time t cash price of this asset, $t = 0, 1, \ldots T$. An *arbitrage opportunity* associated with this asset can be defined just as with a security, except that one must add the stipulation that the position H_1 in this asset must be non-negative. Recalling the economic principles associated with ordinary securities, we realize that an arbitrage opportunity will exist with this asset if we can find some time t and some event $E \in \mathscr{F}_t$ such that

$$A_t(\omega)/B_t(\omega) \leqslant (A_{t+1}(\omega) - c)/B_{t+1}(\omega), \quad \text{all } \omega \in E$$

with this inequality being strict for at least one $\omega \in E$. If event E occurs, then we buy the asset with money borrowed from the bank, and next period we cash out with a net profit that is non-negative in all $\omega \in E$ and strictly positive for at least one such ω. It follows that if there exists a strictly positive probability measure Q such that

$$\text{(4.24)} \qquad A_t/B_t \geqslant E_Q[(A_{t+1} - c)/B_{t+1} | \mathscr{F}_t], \quad t = 0, 1, \ldots, T-1.$$

then this kind of arbitrage opportunity cannot arise.

On the other hand, there is no economic mechanism to prevent inequality (4.24) from being strict for some ω and some t. If this asset $A = \{A_t; T = 0, 1 \ldots, T\}$ is actually a security, then a strict inequality by every strictly positive probability measure would signal an arbitrage opportunity, with the associated trading strategy entailing a short position in the security. But for our asset which cannot be sold short, a strict inequality in (4.24) is of no consequence, because a trading strategy involving a short position is inadmissible and thus not an arbitrage opportunity. Hence I have sketched out an explanation of the following.

(4.25) If the market model includes an asset with cash price $A = \{A_t; T = 0, \ldots, T\}$ which cannot be sold short and which has carrying cost c per unit per period, then there are no arbitrage opportunities if and only if there exists a strictly positive probability measure Q such that all the discounted securities are Q-martingales and such that (4.24) is satisfied.

Now suppose there is a forward price O_t for this asset A, based upon payment at time τ. If this asset is actually a security, then by (4.23) we would have

$$\text{(4.26)} \qquad O_t = \frac{A_t}{E_Q[B_t / B_\tau | \mathscr{F}_t]} + c \sum_{s=t+1}^{\tau} \frac{E_Q[B_t / B_s | \mathscr{F}_t]}{E_Q[B_t / B_\tau | \mathscr{F}_t]}$$

However, if this asset cannot be sold short, then the law of one price does not hold. In particular, if the forward price O_t is strictly less than the right hand side of (4.26), then it would be desirable to buy the asset for time τ delivery at the forward price O_t while simultaneously following the negative of the trading strategy that replicates A_τ. But this cannot be done, because it is not possible to sell this asset short. There is no economic mechanism to prevent the forward price O_t from being strictly less than the right hand side of (4.26).

On the other hand, if the forward price O_t is strictly greater than the right hand side of (4.26), then it would be desirable to sell the asset for time τ delivery at the forward price O_t while simultaneously following the trading strategy that replicates A_τ. This can be done, so this is an arbitrage opportunity. This can be summarized as follows.

(4.27) Suppose there is an asset with cash price $A = \{A_t; t = 0, 1 \ldots, T\}$ which cannot be sold short and which has carrying cost c per unit per period. The time t forward price O_t of this asset, which is received and paid for at time $\tau > t$, satisfies

$$O_t \leqslant \frac{A_t}{E_Q[B_t / B_\tau | \mathscr{F}_t]} + c \sum_{s=t+1}^{\tau} \frac{E_Q[B_t / B_s | \mathscr{F}_t]}{E_Q[B_t / B_\tau | \mathscr{F}_t]}$$

Exercise 4.14 Consider a two-period model with $\Omega = \{\omega_1, \ldots, \omega_4\}$ and an asset with time 0 price $A_0 = 100$.

(a) If the bank account process is deterministic according to $B_t = (1.05)^t$ for $t = 0, 1, 2$ and if the asset can be sold short, then what is the time 0 forward price O_0 of the asset for delivery at time 2?

(b) If the bank account process is random with $B_1 = (1.05)$, $B_2(\omega_1) = B_2(\omega_2) = 1.12$ and $B_2(\omega_3) = B_2(\omega_4) = 1.10$ and if the asset can be sold short, then what is the time 0 forward price O_0 for time 2 delivery of the asset? Provide an expression in terms of $Q(\{\omega_1, \omega_2\})$.

(c) If the bank account process is random as in part (b) and if the asset cannot be sold short, then what is the largest value for the time 0 price O_0 of the asset for time 2 delivery that is consistent with no-arbitrage?

(d) Same as part (c), only now the asset has a carrying cost of \$5 per period.

4.6 Futures

Futures prices closely resemble forward prices in that they both are based on the cash price of a security or asset at a fixed, future point in time. In particular, at time τ both the forward and futures prices for time τ delivery will equal the underlying cash price. However, there are subtle differences which can cause the two prices to be different.

Futures prices are associated with futures contracts which are traded on organized exchanges. For example, a speculator can purchase a contract for the delivery of 5000 bushels of wheat in July. The price is established at the time of the transaction, and then, as additional transactions occur between other buyers and sellers, the value of the speculator's purchased contract will fluctuate in accordance with the ensueing transaction prices. Moreover, the speculator can close out the position at any time before delivery by doing a reverse transaction at the exchange, thereby yielding a net profit based on the difference between the futures prices at the two transaction times. Hence, in the case of futures contracts, there may be arbitrage opportunities which can be 'cashed out' well before the scheduled delivery time.

In contrast, positions in forward contracts must be held until the time of actual delivery. This is because forward contracts are made between two specific individuals or parties, and the position held by one of these parties cannot be closed out before the delivery time by selling the position to a third party. Hence the only arbitrage opportunities that can be present are ones which, once a forward position is established, are held in constant fashion until the delivery time.

Just as with forwards, for every buyer of a futures contract there is a seller, that is, an individual who 'promises to deliver a specified quantity of the asset at a specified future point in time.' This phrase is in quotes, because usually the seller will close out his or her position before the delivery time, and some futures contracts are simply tied to the underlying cash price at the 'delivery' time, with no possibility for an actual delivery to occur.

An important feature of futures contracts is that buyers and sellers must deposit margin or collateral with the exchange. This is to make sure the buyer or seller does not walk away from a losing position. The amount of funds in the trader's margin account will fluctuate continuously in accordance with the fluctuations of subsequent futures prices. For example, if the futures price of a 5000 bushel July wheat contract goes down ten cents a bushel, then a buyer of this contract will have \$500 removed from his or her margin account. In this fashion, cumulative net profits from futures transactions are continuously reflected until the positions are closed out.

Futures traders cannot borrow money and then put the resulting funds in their margin account in order to trade futures contracts. And the exchange will either ask for more funds or close out the trader's futures positions if the margin account becomes depleted. Hence futures traders always have positive wealth, and individuals with zero or negative wealth cannot trade futures contracts.

Futures traders may use securities as collateral in the margin account. For example, a trader could pledge funds earning interest in a bank account. Hence an individual with a security portfolio that has positive wealth can trade futures simply by using part or all of this portfolio as collateral for the margin requirement. This trader is thereby able to trade futures contracts without using any extra funds that are dedicated to the futures contracts. To buy a security you must come up with some cash, but you can buy a futures contract for free!

In order to make these ideas more precise and to see their consequences, we take our usual securities market model and superimpose one or more futures price processes. Let U_t (or $U_j(t)$, $j = 1, \ldots, J$) denote (the notation 'U' comes from *fUtures*) the time t futures price for a security or asset that is delivered at time $\tau \leqslant T$ (or τ_j, respectively). It may be appropriate to stipulate $U_j(\tau_j) = S_j(\tau_j)$ for $j = 1, \ldots, J$, but in some cases the futures market model may contain one or more futures price processes that are not tied to specific securities in this manner. For example, the securities could be stocks and U_t could be the futures price of May corn.

In addition to the usual strategies for trading the securities, let the predictable process $\hat{H}_t$ (or $\hat{H}_j(t)$) represent the position in the futures contract U (or U_j, respectively) held from time $t-1$ to time t. Naturally, if the futures contract U_j expires at time $\tau_j < T$, then $\hat{H}_j(t) = 0$ is required for all $t > \tau_j$. The overall trading strategy will be a predictable, vector-valued process of the form $(H, \hat{H}) = (H_0, H_1, \ldots, H_N, \hat{H}_1, \ldots, \hat{H}_J)$.

At time t, just after any time t transactions, the value of a portfolio is the same as before when there are no futures contracts, namely,

$$H_0(t+1)B_t + \sum_{n=1}^{N} H_n(t+1)S_n(t)$$

However, the portfolio's value V_t just before any time t transactions will be different, for it will equal the money in the bank account and the securities (as with no futures) *plus* the net profit over the last period due to the futures trades. In other words,

$$V_t = H_0(t)B_t + \sum_{n=1}^{N} H_n(t)S_n(t) + \sum_{j=1}^{J} \hat{H}_j(t)\Delta U_j(t), \quad t > 0$$

We let

$$V_0 = H_0(1)B_0 + \sum_{n=1}^{N} H_n(1)S_n(0)$$

denote the usual time 0 value of the portfolio.

A trading strategy $(H, \hat{H})$ in our futures market model will be called *self-financing* if no money is consumed or added to the portfolio from an exogenous source between times 0 and T, that is, if

$$V_t = H_0(t+1)B_t + \sum_{n=1}^{N} H_n(t+1)S_n(t), \quad t = 1, \ldots, T-1$$

Our aim is to derive an explicit relationship for the futures price in terms of its underlying security. We can do this if it is assumed there are no arbitrage opportunities. Since futures traders must have positive wealth, we cannot proceed as with ordinary securities market models and define an arbitrage opportunity as a trading strategy that starts with $V_0 = 0$. But the idea will be the same; we will just shift the starting point to a positive level of initial wealth. A self-financing trading strategy in our futures market model will be called an *arbitrage opportunity* if

(a) $V_T(\omega) \geqslant V_0 B_T(\omega)$, all ω
(b) $V_T(\omega) > V_0 B_T(\omega)$, some ω

Thus an arbitrage opportunity will do no worse than putting all your money in the bank account, and there is the possibility that it will do strictly better.

Just as for ordinary securities market models, we can decompose the multiperiod futures market model into a network of single period models, with one single period model corresponding to each node in the information tree submodel of the filtration. It follows that an arbitrage opportunity will exist for the multiperiod model if and only if there exists an arbitrage opportunity for one or more of the underlying single period models. Hence we can learn about multiperiod futures markets by first studying single period futures markets.

We begin by comparing V_1 with $V_0(B_1/B_0)$:

$$V_1 \geqslant V_0(B_1/B_0) \iff$$

$$H_0(1)B_1 + \sum_{n=1}^{N} H_n(1)S_n(1) + \sum_{j=1}^{J} \hat{H}_j(1)\Delta U_j(1)$$

$$\geqslant \left[H_0(1)B_0 + \sum_{n=1}^{N} H_n(1)S_n(0)\right](B_1/B_0) \iff$$

$$\sum_{n=1}^{N} H_n(1)S_n(1) + \sum_{j=1}^{J} \hat{H}_j(1)\Delta U_j(1) \geqslant B_1 \sum_{n=1}^{N} H_n(1)S_n(0)/B_0 \iff$$

$$\sum_{n=1}^{N} H_n(1)S_n^*(1) + \sum_{j=1}^{J} \hat{H}_j(1)\Delta U_j(1)/B_1 \geqslant \sum_{n=1}^{N} H_n(1)S_n^*(0) \iff$$

(4.28) $$\sum_{n=1}^{N} H_n(1)\Delta S_n^*(1) + \sum_{j=1}^{J} \hat{H}_j(1)\Delta U_j(1)/B_1 \geqslant 0$$

Note the first inequality is strict for one or more states ω if and only if inequality (4.28) is strict. Thus $(H, \hat{H})$ is an arbitrage opportunity for this single period model if and only if (4.28) holds, with the inequality being strict for one or more $\omega \in \Omega$.

Looking at (4.28) we see that the quantity $\Delta U_j(1)/B_1$ plays the same kind of role as the quantity $\Delta S_n^*(1)$. Knowing what we do for conventional single period models, we conclude the following.

(4.29) There are no arbitrage opportunities in the single period futures market model if and only if there exists a strictly positive probability measure Q such that $S_n^*(0) = E_Q[S_n^*(1)]$, $n = 1, \ldots, N$ and that $U_j(0) = E_Q[U_j(1)/B_1]/E_Q\ [1/B_1]$, $j = 1, \ldots, J$. If the interest rate $r = (B_1 - B_0)/B_0$ is constant, then this last equation simplifies to $U_j(0) = E_Q U_j(1)$.

Now (4.29) can be viewed as a requirement for the risk neutral probability measure. Given the bank account, security, and futures prices, it gives the requirement which Q must satisfy for there to be no arbitrage opportunities. Notice the requirement that $U_j(\tau_J) = S_j(\tau_j)$ for some time τ_j was not used; we only used the futures exchange's trading rules. Hence (4.29) can be applied to situations where expiration times τ exceed one (or even to situations where a futures price is based on an asset that is not a security or on a security that is not part of the securities model).

Alternatively, one could start with an arbitrage free securities market model and wish to add futures prices based on these securities. In the case of the single period model, one would know Q (this comes from the original securities market model) and the fact that $U_j(1) = S_j(1)$, and so one would use (4.29) and the fact that $S_j(0) = E_Q[S_j(1)/B_1]$ to conclude

(4.30) $$U_j(0) = \frac{S_j(0)}{E_Q[1/B_1]}$$

Example 4.11 Consider the single period model with $N = 1$, $K = 2$, $S_0 = 5$, $S_1(\omega_1) = U_1(\omega_1) = 8$, $S_1(\omega_2) = U_1(\omega_2) = 4$, and the interest

rate constant with $0 \leqslant r < 3/5$. The risk neutral probability measure is computed from $E_Q[\Delta S^*] = 0$ to be $Q(\omega_1) = (1+5r)/4$ and $Q(\omega_2) = (3-5r)/4$. Hence by (4.29) the futures market model is arbitrage free if and only if $U_0 = E_Q[S_1] = 5(1+r) = 5B_1$. Of course, this is the same as (4.30) and is consistent with the fact that

$$\textbf{(4.31)} \qquad V_1 = H_0 B_1 + H_1 S_1 + \hat{H}\Delta U = (V_0 - 5H_1)B_1 + H_1 S_1 + \hat{H}\Delta U$$
$$= \begin{cases} V_0 B_1(\omega_1) + H_1(8 - 5B_1(\omega_1)) + \hat{H}\Delta U(\omega_1), & \omega = \omega_1 \\ V_0 B_1(\omega_2) + H_1(4 - 5B_1(\omega_2)) + \hat{H}\Delta U(\omega_2), & \omega = \omega_2 \end{cases}$$
$$= \begin{cases} V_0(1+r) + H_1(3-5r) + \hat{H}(3-5r), & \omega = \omega_1 \\ V_0(1+r) + H_1(-1-5r) + \hat{H}(-1-5r), & \omega = \omega_2 \end{cases}$$

The point here is that with $U_0 = 5(1+r)$ there is no way to choose H_1 and $\hat{H}$ so as to make $V_1 \geqslant V_0(1+r)$ and $V_1 \neq V_0(1+r)$. On the other hand, with U_0 taking any other value, the vector ΔU will not be a scalar multiple of the vector $S_1 - S_0 B_1$, and one can choose H_1 and $\hat{H}$ so as to be an arbitrage opportunity.

More generally, if U_1 is arbitrary instead of being equal to S_1, then (4.29) implies $U_0 = U_1(\omega_1)(1+5r)/4 + U_1(\omega_2)(3-5r)/4$. A little algebra verifies that the vector ΔU_1 is a scalar multiple of the vector $S_1 - S_0 B_1$, which is consistent with no arbitrage.

Now return to the case where $U_1 = S_1$ and suppose the interest rate r is random. The formula $U_0 = E_Q[S_1]$ in (4.29) will not apply, and instead we must use the more general $U_0 = E_Q[S_1/B_1]/E_Q[1/B_1]$ or (4.30). For example, with $r(\omega_1) = 1/16$ and $r(\omega_2) = 1/8$, the risk neutral probability measure is computed to be $Q(\omega_1) = 221/608$ and $Q(\omega_2) = 387/608$. It follows from (4.30) that $U_0 = 5\ 35/69$, which is different than $E_Q S_1 = 5\ 69/152$. With $U_0 = 5\ 35/69$ one has $\Delta U(\omega_1) = 172/69$ and $\Delta U(\omega_2) = -104/69$, so the first part of (4.31) implies

$$V_1 = \begin{cases} \frac{17}{16} V_0 + \frac{43}{16} H_1 + \frac{172}{69}\hat{H}, & \omega = \omega_1 \\ \frac{9}{8} V_0 - \frac{26}{16} H_1 - \frac{104}{69}\hat{H}, & \omega = \omega_2 \end{cases}$$

Some more algebra verifies that the vector ΔU is a scalar multiple of $S_1 - S_0 B_1$, so there is no way of choosing H_1 and $\hat{H}$ so as to be an arbitrage opportunity. Similar results hold when B_1 is random and $U_1 \neq S_1$.

In the preceding example, when $U_1 = S_1$ the futures price U_0 came out equal to the forward price $O_0 = S_0/E_Q[1/B_1]$. Looking at (4.30) one sees that this will hold in all single period models, even if B_1 is random, when $U_1 = S_1$. This is not surprising, since in a single period model the trading strategies available for replicating mispriced forwards and futures are identical. But in a multiperiod setting only 'buy-and-hold' and 'short-and-hold' strategies can be used to replicate mispriced forwards, and so, as we shall see, the futures and forward prices can be different.

We are now ready to analyze the multiperiod futures market model. Consider the single period model associated with an arbitrary time t and event A in $\mathscr{P}_t$, the cell in the partition corresponding to $\mathscr{F}_t$. As stated earlier, the multiperiod futures market model has no arbitrage opportunities if and only if none of the underlying single period models has any arbitrage opportunities. Our arbitrary single period model has an arbitrage opportunity if and only if trading positions can be taken at time t such that $V_{t+1} 1_A \geqslant 1_A V_t B_{t+1}/B_t$, with this inequality being strict for at least one $\omega \in A$. Proceeding in an analogous fashion through the implications that led to (4.28), we see that the single period model has an arbitrage opportunity if and only if

$$\sum_{n=1}^{N} H_n(t+1)\Delta S_n^*(t+1) + \sum_{j=1}^{J} \hat{H}_j(t+1)\Delta U_j(t+1)/B_{t+1} \geqslant 0, \text{ all } \omega \in A$$

with this inequality being strict for at least one $\omega \in A$. It follows as in (4.29) that this single period model has no arbitrage opportunities if and only if there is a conditional risk neutral probability measure $Q(t, A) > 0$ such that $E_{Q(t,A)}[\Delta S_n^*(t+1)] = 0$ for all n and $E_{Q(t,A)}[\Delta U_j(t+1)/B_{t+1}] = 0$ for all j. Finally, and just as with conventional, multiperiod securities markets models (see section 3.4), we can paste all the single period models together and conclude that the multiperiod futures market model has no arbitrage opportunities if and only if there exists a strictly positive probability measure Q such that each S_n^* is a Q-martingale and

$$\textbf{(4.32)} \qquad E_Q[\Delta U_j(t+1)/B_{t+1} | \mathscr{F}_t] = 0, \quad \text{all } t \geqslant 0 \text{ and } j = 1, \ldots, J$$

There is just one problem: condition (4.32) is rather distasteful, being unattractive and somewhat difficult to work with. It turns out that little will be lost and considerable simplification will be gained if it is assumed that the bank account process B is predictable (in the single period model, this is the same as requiring B_1 to be a constant), for then (4.32) reduces to the much nicer requirement that each (undiscounted) futures process is a martingale under Q. This is the approach we will take; the results are summarized in the following.

(4.33) Suppose the bank account process is predictable. Then there are no arbitrage opportunities in the multiperiod futures market model if and only if there exists a strictly positive probability measure Q such that each S_n^*, $n = 1, \ldots, N$, and each U_j, $j = 1, \ldots, J$, is a martingale under Q.

Now suppose U is the price of a futures contract based on the delivery of security S at time τ, so $U_\tau = S_\tau$. Then (4.33) implies $U_t = E_Q[S_\tau | \mathscr{F}_t]$ for all $t = 0, \ldots, \tau$. Meanwhile, consider the forward price O_t for delivery of the

same security S at the same time τ. If this security pays no dividend, then (4.22) says

$$O_t = \frac{S_t}{E_Q[B_t/B_\tau|\mathscr{F}_t]} = \frac{E_Q[S_\tau B_t/B_\tau|\mathscr{F}_t]}{E_Q[B_t/B_\tau|\mathscr{F}_t]} = \frac{E_Q[S_\tau/B_\tau|\mathscr{F}_t]}{E_Q[1/B_\tau|\mathscr{F}_t]}$$

Comparing this with $U_t = E_Q[S_\tau|\mathscr{F}_t]$, we realize the following.

(4.34) When the bank account process B is deterministic, the forward and futures prices for time τ delivery of the same security coincide, whereas if B is random (and predictable), then these forward and futures price processes can be different.

This same conclusion holds when the underlying security pays a dividend, as can be verified by using (4.23) and (3.28).

Example 4.1 (continued) Suppose $B_1 = 1$ and

$$B_2(\omega) = \begin{cases} 17/16, & \omega = \omega_1, \omega_2 \\ 9/8, & \omega = \omega_3, \omega_4 \end{cases}$$

The unique probability measure such that S^* is a martingale is easily computed to be $Q(\omega_1) = 5/24$, $Q(\omega_2) = 1/24$, and $Q(\omega_3) = Q(\omega_4) = 3/8$. Hence for the futures price process U satisfying $U_2 = S_2$ one has $U_0 = E_Q[S_2] = 5\frac{1}{2}$ and

$$U_1 = E_Q[S_2|\mathscr{F}_1] = \begin{cases} 8\frac{1}{2}, & \omega = \omega_1, \omega_2 \\ 4\frac{1}{2} & \omega = \omega_3, \omega_4 \end{cases}$$

Meanwhile, the time 0 forward price for time 2 delivery of this security is given by $O_0 = S_0/E_Q[1/B_2] = 5/(46/51) = 525/46$.

Now turn to the topic of options on futures. For example, suppose contingent claim X is defined by $X = g(U_s)$ for some expiration date s satisfying $0 < s < \tau$. More generally, X is an $\mathscr{F}_s$ measurable random variable with s smaller than all the delivery times of the various futures contracts associated with the model. The objectives are to determine the price of this European option and to derive a trading strategy which replicates this option, that is, a self- financing trading strategy $(H, \hat{H})$ satisfying $V_s = X$, where, as above,

$$V_t = H_0(t)B_t + \sum_{n=1}^{N} H_n(t)S_n(t) + \sum_{j=1}^{J} \hat{H}_j(t)\Delta U_j(t), \quad t > 0$$

As usual, an option X will be called attainable or marketable if it can be replicated. And if X can be replicated by the trading strategy $(H, \hat{H})$, then the corresponding portfolio value V_t must be the time t price of X for all $t \leqslant s$.

In order to obtain nice results, from now on it will be assumed that the bank account process B is predictable and that there exists a risk neutral probability measure Q for this futures market model. It follows from this assumption and (4.33) that:

(4.35) The discounted value process $V^* \equiv V/B$ is a Q-martingale.

Now X is attainable if and only if $X/B_s = V^*_s$ for some self-financing trading strategy, so just as with ordinary security market models (4.35) implies

(4.36) If $X \in \mathscr{F}_s$ is an attainable contingent claim in the futures market model, then its time t price is $V_t = E_Q[XB_t/B_s|\mathscr{F}_t]$ for all $t \leqslant s$.

It is worth pointing out that results like (4.35) and (4.36) hold even if the bank account process B is not predictable, but you must be careful what you mean by the risk neutral probability measure in such a case.

This valuation formula for options on futures looks exactly the same as ordinary European options on securities, but there is a subtle difference. In practical applications it is common to work with a model consisting of the bank account and just one risky price process (chosen to be, of course, the price underlying the option). If this underlying risky price is a security, then the martingale measure Q will be such that the *discounted* risky price process is a martingale. On the other hand, if the underlying risky price is a futures price, then the martingale measure Q must be such that the *undiscounted* risky price process is a martingale. Thus starting with the same model for the risky price process, you can get two different prices for the same option on this price process, depending on whether it is a security or a futures price.

Example 4.12 Consider a futures market model that consists of a single risky security, namely, a futures price process U that is governed by the binomial model of section 3.5. In particular, for parameters $0 < d < 1 < u$ and $0 < p < 1$ and initial price U_0,

$$U_t = U_0 u^{N_t} d^{t-N_t} \quad t = 0, 1, \ldots$$

where N is a binomial process with parameter p. As usual, the interest rate $r \geqslant 0$ is constant.

The risk neutral probability measure is easy to obtain. We want U to be a martingale, so this is simply the same as before, only with the interest rate $r = 0$. In particular, the probability q of an 'up' move should be taken to be

$$q = \frac{1-d}{u-d}$$

Everything else should be the same as before, even if the true interest rate r is strictly positive. In particular, formula (3.32) still holds for the probability distribution of U_t, except that we now use the new value of q.

Now consider the European call option $X = (U_s - e)^+$. To compute its time 0 price we proceed in exactly the same way as in section 4.2,

only using the new value of q. In particular, we define $\hat{n}$ to be the smallest non-negative integer n such that

$$n > \frac{\log(e/(U_0 d^s))}{\log(u/d)}$$

If $\hat{n} > s$, then U_0 is so far out of the money that there is no chance of finishing in the money (that is, with $U_s > e$)), in which case the time 0 price of the option is $V_0 = 0$. On the other hand, if $\hat{n} \leqslant s$, then

$$V_0 = U_0 \sum_{n=\hat{n}}^{s} \binom{s}{n} \hat{q}^n (1-\hat{q})^{s-n} - \frac{e}{(1+r)^s} \sum_{n=\hat{n}}^{s} \binom{s}{n} q^n (1-q)^{s-n}$$

where $q = (1-d)/(u-d)$ and $\hat{q} = qu/(1+r) = (1-d)u/[(u-d)(1+r)]$. This formula has exactly the same form as for a call option on a security, only now $q \neq (1+r-d)/(u-d)$.

Exercise 4.15 Verify that (4.34) holds when the underlying security pays a dividend.

5 Optimal Consumption and Investment Problems

5.1 Optimal Portfolios and Dynamic Programming

The purpose of this chapter is to study the multiperiod generalizations of the single period consumption and investment problems that were introduced in chapter 2. I begin by investigating the basic optimal portfolio problem, where the objective is to maximize the expected utility of time T wealth and where there is no consumption before time T.

A *utility function* $u : \mathbb{R} \times \Omega \to \mathbb{R}$ is specified, with $u(w, \omega)$ representing the utility of wealth w at time T when $\omega \in \Omega$ is the state of the world. It will be assumed that $w \to u(w, \omega)$ is differentiable, concave, and strictly increasing for each $\omega \in \Omega$. Usually u will be independent of ω.

An initial wealth v is specified. The investor can choose any self-financing trading strategy H consistent with this initial wealth. The measure of performance of any such H will be the expected utility of terminal wealth, that is,

$$Eu(V_T) = \sum_{\omega \in \Omega} P(\omega)\, u(V_T(\omega), \omega)$$

The investor is therefore interested in solving the following optimal portfolio problem:

$$\text{(5.1)} \qquad \begin{aligned} &\text{maximize } Eu(V_T) \\ &\text{subject to } V_0 = v \\ &H \in \mathbb{H} \end{aligned}$$

where $\mathbb{H}$ denotes the set of all self-financing trading strategies.

Keeping in mind that trading strategies must be predictable, we see that (5.1) actually involves three kinds of constraints. But two of these three kinds can easily be set aside. Since $V_T = B_T V_T^*$ and $V_T^* = V_0^* + G_T^*$, it follows that (5.1) is equivalent to

(5.2)
$$\begin{aligned}&\text{maximize} \quad Eu(B_T\{v + G_T^*\})\\ &\text{subject to} \quad (H_1, \ldots, H_N) \in \mathbb{H}_p\end{aligned}$$

where $\mathbb{H}_p$ denotes the set of all predictable processes that take values in $\mathbb{R}^N$. Thus if $(\hat{H}_1, \ldots, \hat{H}_N)$ is a solution of (5.2), then it is a simple matter to choose $\hat{H}_0$ so that $\hat{H} = (\hat{H}_0, \hat{H}_1, \ldots, \hat{H}_N)$ is self-financing and $V_0 = v$, thereby giving an optimal solution of (5.1).

Problem (5.2) can be solved with several approaches. One is to use standard calculus and optimization theory, keeping in mind the predictability constraint.

Example 5.1 Suppose $T = 2, K = 4, N = 1$, the interest rate r is a constant satisfying $0 \leqslant r < 0.125$, the filtration is the one generated by the risky security, and the price process for the risky security and the probability measure are as follows:

ω	$S_0(\omega)$	$S_1(\omega)$	$S_2(\omega)$	$P(\omega)$
ω_1	5	8	9	1/4
ω_2	5	8	6	1/4
ω_3	5	4	6	1/4
ω_4	5	4	3	1/4

In addition, suppose the investor has an exponential utility function: $u(w) = 1 - \exp\{-w\}$. In view of the predictability requirement, the strategy H_1 for trading the risky security entails the specification of three scalar values: the position, denoted H^5, carried forward from time 0 when the price $S_0 = 5$, the position, denoted H^8, carried forward from time 1 when the price $S_1 = 8$, and the position, denoted H^4, carried forward from time 1 when the price $S_1 = 4$. Hence the objective function in (5.2) can be written as

$$\begin{aligned}
Eu(B_2\{v + G_2^*\}) &= 1 - E\exp\{-(1+r)^2[v + H_1(1)\Delta S_1^* + H_1(2)\Delta S_2^*]\}\\
&= 1 - \frac{1}{4}\left(\exp\left\{-(1+r)^2\left[v + H^5\frac{3-5r}{1+r} + H^8\frac{1-8r}{(1+r)^2}\right]\right\}\right.\\
&\quad + \exp\left\{-(1+r)^2\left[v + H^5\frac{3-5r}{1+r} + H^8\frac{-2-8r}{(1+r)^2}\right]\right\}\\
&\quad + \exp\left\{-(1+r)^2\left[v + H^5\frac{-1-5r}{1+r} + H^4\frac{2-4r}{(1+r)^2}\right]\right\}\\
&\quad \left. + \exp\left\{-(1+r)^2\left[v + H^5\frac{-1-5r}{1+r} + H^4\frac{-1-4r}{(1+r)^2}\right]\right\}\right)
\end{aligned}$$

Setting the partial derivative with respect to each of the three variables equal to zero leads to the following three equations:

(5.3)

$$\begin{aligned}
&\exp\left\{-(1+r)^2\left[v + H^5\frac{3-5r}{1+r} + H^8\frac{1-8r}{(1+r)^2}\right]\right\}\\
&\quad = \frac{2+8r}{1-8r}\exp\left\{-(1+r)^2\left[v + H^5\frac{3-5r}{1+r} + H^8\frac{-2-8r}{(1+r)^2}\right]\right\}\\
&\exp\left\{(1+r)^2\left[v + H^5\frac{-1-5r}{1+r} + H^4\frac{2-4r}{(1+r)^2}\right]\right\}\\
&\quad = \frac{(3-5r)\,(1+4r)}{(1+5r)\,(1-8r)}\exp\left\{-(1+r)^2\left[v + H^5\frac{3-5r}{1+r} + H^8\frac{-2-8r}{(1+r)^2}\right]\right\}\\
&\exp\left\{-(1+r)^2\left[v + H^5\frac{-1-5r}{1+r} + H^4\frac{-1-4r}{(1+r)^2}\right]\right\}\\
&\quad = \frac{(3-5r)\,(2-4r)}{(1+5r)\,(1-8r)}\exp\left\{-(1+r)^2\left[v + H^5\frac{3-5r}{1+r} + H^8\frac{-2-8r}{(1+r)^2}\right]\right\}
\end{aligned}$$

Taking logs, one obtains a system of three linear equations; from this the final solution is easily obtained:

$$\begin{aligned}
H^5 &= \frac{3\ln(3-5r) + (2-4r)\ln(2-4r) + (1+4r)\ln(1+4r)}{12(1+r)}\\
&\quad - \frac{3\ln(1+5r) + (2+8r)\ln(2+8r) + (1-8r)\ln(1-8r)}{12(1+r)}
\end{aligned}$$

$$H^8 = -\frac{1}{3}\ln\left(\frac{2+8r}{1-8r}\right) \qquad H^4 = \frac{1}{3}\ln\left(\frac{2-4r}{1+4r}\right)$$

It remains to compute H_0, the strategy for trading the bank account. Clearly $H_0(1) = v - 5H^5$. In states ω_1 and ω_2 the value of the portfolio will be $V_1 = (v - 5H^5)(1+r) + 8H^5$, so setting this equal to $H_0(2)(1+r) + 8H^8$ gives $H_0(2) = (v - 5H^5) + 8(H^5 - H^8)/(1+r)$. Similarly, one computes $H_0(2) = (v - 5H^5) + 8(H^5 - H^4)/(1+r)$ for states ω_3 and ω_4.

As seen in the case of single period models, if there exists an arbitrage opportunity, then there cannot exist an optimal solution to portfolio problems (5.1) or (5.2). In other words, if (5.1) or (5.2) has a solution, then there are no arbitrage opportunities, in which case there must exist a risk neutral probability measure. Indeed, principle (2.6) for single period models generalizes to the following:

(5.4) If (H,V) is a solution of the optimal portfolio problem (5.1) or (5.2), then a risk neutral probability measure is defined by

$$Q(\omega) = \frac{P(\omega)B_T u'(V_T(\omega),\omega)}{E[B_T u'(V_T)]}, \quad \omega \in \Omega$$

where u' denotes the partial derivative with respect to the first argument.

To see this, consider an arbitrary time t, arbitrary security n, and arbitrary event A in $\mathcal{P}_{t-1}$, the partition corresponding to $\mathcal{F}_{t-1}$. Corresponding to A is $H_n(t)1_A$, the position in security n that is carried forward from time $t-1$ when event A occurs. The first order necessary condition corresponding to this scalar-valued variable is

$$\sum_{\omega \in A} P(\omega)u'(B_T(\omega)\{v + G_T^*(\omega)\},\omega)B_T(\omega)\Delta S_n^*(t,\omega) = 0$$

This is true for all $A \in \mathcal{P}_{t-1}$, so

$$E[u'(B_T\{v + G_T^*\})B_T\Delta S_n^*(t)|\mathcal{F}_{t-1}] = 0$$

Hence $E_Q[\Delta S_n^*(t)|\mathcal{F}_{t-1}] = 0$ if Q is defined as in (5.4), since $V_T = B_T\{v + G_T^*\}$.

Example 5.1 (continued) The three equations in (5.3) correspond respectively to

$$u'(V_2(\omega_1)) = \frac{2+8r}{1-8r}u'(V_2(\omega_2))$$

$$u'(V_2(\omega_3)) = \frac{(3-5r)(1+4r)}{(1+5r)(1-8r)}u'(V_2(\omega_2))$$

$$u'(V_2(\omega_4)) = \frac{(3-5r)(2-4r)}{(1+5r)(1-8r)}u'(V_2(\omega_2))$$

Hence (5.4) implies

$$Q(\omega_1) = \frac{(1+5r)(2+8r)}{12} \qquad Q(\omega_2) = \frac{(1+5r)(1-8r)}{12}$$

$$Q(\omega_3) = \frac{(3-5r)(1+4r)}{12} \qquad Q(\omega_4) = \frac{(3-5r)(2-4r)}{12}$$

Conventional methods verify that this is the unique risk neutral probability measure.

It should be apparent that the approach illustrated in example 5.1 for computing the optimal trading strategy may be impractical for large problems. With N equations and N variables for each node of the underlying information tree, the resulting system of equations may be too large to solve.

But an alternative approach called dynamic programming may reduce these computational difficulties.

The dynamic programming idea was already introduced in connection with computing the value of American options. The idea is to realize that when faced with a sequence of decisions, the optimal decision to make now should be consistent with the intention to act optimally in all future periods. In other words, if you know the optimal strategy starting at time $t+1$, then the determination of the optimal strategy starting at time t can be reduced to one or more one-period problems. It follows that a multiperiod decision problem can be solved by solving a sequence of one-period problems. You work backwards in time, first computing the optimal decisions with one period to go, then computing the optimal decisions with two periods to go, and so forth.

In order to implement this procedure in the case of our optimal portfolio problem, it is necessary to keep track of the *optimal value process* $U_t(w), t = 0, \ldots, T$. Here $U_t(w)$ equals the maximum (over all self-financing trading strategies) expected utility of time T wealth given it is now time t, the time t wealth is w, and the time t history is $\mathcal{F}_t$. Hence $U_t(w)$ will be an $\mathcal{F}_t$ measurable random variable.

The value of $U_t(w)$ when $t = T$ is clear: it coincides with the utility function, that is,

$$U_T(w) = u(w,\omega)$$

It turns out that for $t < T$ the value of $U_t(w)$ satisfies the important *dynamic programming functional equation*:

$$U_t(w) = \max_{H \in \mathcal{F}_t} E[U_{t+1}(B_{t+1}\{w/B_t + H \cdot \Delta S_{t+1}^*\})|\mathcal{F}_t] \tag{5.5}$$

Here H, the decision variable for period t, is an N-dimensional random variable that is required to be $\mathcal{F}_t$ measurable. The value of H that maximizes this expression will turn out to be the vector of optimal positions in the risky securities carried forward from time t, given the history $\mathcal{F}_t$. The notation $H \cdot \Delta S_{t+1}^*$ represents the inner product, that is, $H \cdot \Delta S_{t+1}^* = H_1\Delta S_1^*(t+1) + \ldots + H_N\Delta S_N^*(t+1)$. This equals the discounted gain from time t to time $t+1$. Notice that the argument of U_{t+1} in (5.5) equals the time $t+1$ wealth if the time t wealth is w, H gives the positions in the risky securities, and $H_0(t+1)$, the position in the bank account, is chosen in a self-financing manner.

The dynamic programming equation (5.5) can be used to compute an optimal solution to the portfolio problem (5.1) or (5.2) by computing the optimal value functions $U_t(w)$ in a recursive manner. First compute $U_{T-1}(w)$, then compute $U_{T-2}(w)$, and so forth. Along the way keep track of the maximizing values of H, for these will form the components of the optimal trading strategy. When completed, $U_0(w)$ will equal the optimal value of the objective function in (5.1) or (5.2) given $w = v$. Thus the dynamic programming method provides a bonus: you have a

solution for all possible values of the initial wealth $w = v$, not just a specific value.

Example 5.1 (continued) Taking $t = 1$ and either ω_1 or ω_2, the right hand side of (5.5) becomes

$$\max_h E[1 - \exp(-(1+r)^2\{w/(1+r) + h\Delta S_2^*\})|S_1 = 8]$$

$$= \max_h (1 - \frac{1}{2}\exp\{-(1+r)w - (1-8r)h\} - \frac{1}{2}\exp\{-(1+r)w + (2+8r)h\})$$

The decision variable h here is a scalar. Computing the derivative of this argument with respect to h and setting this equal to zero soon leads to the maximizing value of h, namely,

$$h = -\frac{1}{3}\ln\left(\frac{2+8r}{1-8r}\right)$$

Substituting this back into the right hand side of (5.5) yields

$$U_1(w) = 1 - \frac{1}{2}\exp\{-(1+r)w\}\left\{\left(\frac{2+8r}{1-8r}\right)^{(1-8r)/3} + \left(\frac{2+8r}{1-8r}\right)^{-(2+8r)/3}\right\}$$

$$= 1 - \frac{3}{2}(2+8r)^{-(2+8r)/3}(1-8r)^{-(1-8r)/3}\exp\{-(1+r)w\}$$

for ω_1 and ω_2.

In a similar fashion, taking $t = 1$ and either ω_3 or ω_4, the right hand side of (5.5) becomes

$$\max_h E[1 - \exp(-(1+r)^2\{w/(1+r) + h\Delta S_2^*\})|S_1 = 4]$$

$$= \max_h (1 - \frac{1}{2}\exp\{-(1+r)w - (2-4r)h\} - \frac{1}{2}\exp\{-(1+r)w + (1+4r)h\})$$

so $\ln[(2-4r)/(1+4r)]/3$ is the maximizing value of h and

$$U_1(w) = 1 - \frac{1}{2}\exp\{-(1+r)w\}\left\{\left(\frac{2-4r}{1+4r}\right)^{(1+4r)/3} + \left(\frac{2-4r}{1+4r}\right)^{-(2-4r)/3}\right\}$$

$$= 1 - \frac{3}{2}(2-4r)^{-(2-4r)/3}(1+4r)^{-(1+4r)/3}\exp\{-(1+r)w\}$$

for ω_3 and ω_4.

We are now ready to do the dynamic programming iteration and compute $U_0(w)$. Denote

$$f(r,\omega) = \begin{cases} \frac{3}{2}(2+8r)^{-(2+8r)/3}(1-8r)^{-(1-8r)/3}, & \omega = \omega_1, \omega_2 \\ \frac{3}{2}(2-4r)^{-(2-4r)/3}(1+4r)^{-(1+4r)/3}, & \omega = \omega_3, \omega_4 \end{cases}$$

so $U_1(w)$ can be written concisely as

$$U_1(w) = 1 - f(r,\omega)\exp\{-(1+r)w\}$$

Equation (5.5) becomes

$$U_0(w) = \max_h E[1 - f(r,\omega)\exp\{-(1+r)[(1+r)\{w + h\Delta S_1^*\}]\}]$$

$$= \max_h (1 - \frac{1}{2}f(r,\omega_1)\exp\{-(1+r)^2 w - (1+r)(3-5r)h\}$$

$$- \frac{1}{2}f(r,\omega_3)\exp\{-(1+r)^2 w + (1+r)(1+5r)h\})$$

Setting the derivative with respect to h equal to zero, one is eventually led to the maximizing value of h:

$$h = \frac{3\ln(3-5r) + (2-4r)\ln(2-4r) + (1+4r)\ln(1+4r)}{12(1+r)} - \frac{3\ln(1+5r) + (2+8r)\ln(2+8r) + (1-8r)\ln(1-8r)}{12(1+r)}$$

Substituting this back into the right hand side of (5.5) enables one to obtain an expression for $U_0(w)$. However, the algebraic details are excessive, so an actual formula will not be provided.

In summary, the dynamic programming approach can, in principle, be used to solve the optimal portfolio problem. It provides solutions for many cases where a conventional approach, based on the first-order necessary conditions, is inadequate. On the other hand, there are many practical situations where the computational difficulties are formidable, if not unsurmountable. Fortunately, the risk neutral computational approach, which is the subject of the next section, can often overcome these computational difficulties.

Exercise 5.1 Use dynamic programming to compute the optimal trading strategy as a function of the initial wealth v for the securities market in example 5.1. Use log utility, that is, $u(w) = \ln(w)$, and assume the interest rate r is an arbitrary constant. Compute the optimal objective value as a function of the parameters r and v. (Hints: try the special case $r = 0$ first; show the optimal time 0 position in the risky security is $(1+r)(1-5r)v/[(1+5r)(3-5r)]$.)

Exercise 5.2 Use dynamic programming to show that with exponential utility (that is, $u(w) = a - (b/c)\exp\{-cw\}$, where a, $b > 0$ and $c > 0$ are

scalar parameters), with a general securities model, and with a predictable bank account process, that the optimal position in the securities is always independent of the current wealth. What happens if you drop the predictability requirement? Give a proof or use the security in example 5.1 to provide a counter-example.

Exercise 5.3 Suppose there is a single risky security that follows the binomial model with parameters $0 < d < 1+r < u$ and $0 < p < 1$. The interest rate $r \geqslant 0$ is constant. The initial price S_0, the initial wealth v, and the time horizon T are arbitrary. Use dynamic programming to compute the optimal trading strategy for the following utility functions:

(a) $u(w) = 1 - \exp\{-w\}$ (Hint: show by induction that $U_t(w))$ is of the form $1 - k_t \exp\{-(1+r)^{T-t}w\}$, where k_t is a constant.)
(b) $u(w) = \ln(w)$ (Hint: show by induction that $U_t(w)$ is of the form $\ln(w) + k_t$, where k_t is a constant.)

Exercise 5.4 For the model in example 5.1 with constant interest rate $r = 0$, compute the optimal trading strategy for the indicated utility function, using both the standard optimization approach (i.e., set the three partials equal to zero) and dynamic programming. Verify that your answers are the same.

(a) $u(w) = -w^{-1}$ (isoelastic utility)
(b) $u(w) = \beta w - \frac{1}{2}w^2$ (quadratic utility).

5.2 Optimal Portfolios and Martingale Methods

The risk neutral computational approach for solving multiperiod optimal portfolio problems is pretty much the same as for solving single period problems. Given problem (5.1) or (5.2), the first step is to identify $\mathbb{W}_v$, the set of all the *attainable wealths*. This is $\mathbb{W}_v = \{w \in \mathbb{R}^K : W = V_T$ for some self-financing H with $V_0 = v\}$, the set of all the time T contingent claims that can be generated by some self-financing trading strategy starting with initial wealth v. If the model is complete, this set is simply

$$\mathbb{W}_v = \{W \in \mathbb{R}^K : E_Q[W/B_T] = v\} \tag{5.6}$$

If the model is not complete, then the specification of $\mathbb{W}_v$ is more complex and will be discussed in a later section.

The second step is to solve the subproblem:

$$\begin{aligned} &\text{maximize} \quad && Eu(W) \\ &\text{subject to} \quad && W \in \mathbb{W}_v \end{aligned} \tag{5.7}$$

If the model is complete, then this problem can be solved with a Lagrange multiplier technique, as will be explained in a moment. Finally, having obtained the optimal solution $\hat{W}$, say, the third step is to compute the trading strategy H which generates $\hat{W}$, doing this exactly as one would compute the trading strategy that replicates a contingent claim.

Throughout this section it will be assumed that the model is complete, so the only step requiring explanation is the second: solving subproblem (5.7). An efficient procedure will now be described, and examples will be provided.

Actually, the second step is very little different from what was done for single period models. In view of (5.6), we want to solve (5.7) by introducing the Lagrange multiplier λ and then solving

$$\text{maximize} \quad Eu(W) - \lambda E_Q[W/B_T] \tag{5.8}$$

This is an unconstrained problem with the variable $W \in \mathbb{R}^K$. Introducing the state price density $L = Q/P$, the objective function in (5.8) can be rewritten as

$$E[u(W) - \lambda LW/B_T] = \sum_{\omega \in \Omega} P(\omega)[u(W(\omega)) - \lambda L(\omega) W(\omega)/B_T(\omega)]$$

If W maximizes this expression, then the necessary conditions must be satisfied, giving rise to one equation for each $\omega \in \Omega$:

$$u'(W(\omega)) = \lambda L(\omega)/B_T(\omega), \quad \text{all } \omega \in \Omega$$

(it is now being assumed that the utility function $u : \mathbb{R} \to \mathbb{R}$ is a function only of wealth, independent of the state $\omega \in \Omega$). This is equivalent to

$$W(\omega) = I(\lambda L(\omega)/B_T(\omega)), \quad \text{all } \omega \in \Omega \tag{5.9}$$

where I denotes the inverse function corresponding to u'.

All that remains is to determine the correct value of λ. This is simply the value such that $v = E_Q[W/B_T]$ is satisfied when (5.9) is substituted for W. In other words, λ should be chosen to satisfy

$$E_Q[I(\lambda L/B_T)/B_T] = v \tag{5.10}$$

The inverse function I is normally decreasing with a range that includes $(0, \infty)$, so normally a solution λ to (5.10) will exist for any $v > 0$. Hence the solution of subproblem (5.7) is really no different from what it was for single period models, except that now we discount by B_T rather than B_1. In view of the single period results in section 2.2, we immediately have the following examples.

Example 5.2 (exponential utility) The exponential utility function of example 2.2 can easily be generalized to be of the form $u(w) = a - bc \exp\{-w/c\}$, where a, b, and c are scalar parameters with $b > 0$ and $c > 0$. This gives the optimal attainable wealth

$$W = \frac{v + cE[(L/B_T)\ln(L/B_T)]}{E[L/B_T]} - c\ln(L/B_T)$$

and the optimal objective value

$$Eu(W) = a - bcE[L/B_T] \exp\left\{\frac{-v/c - E[(L/B_T)\ln(L/B_T)]}{E[L/B_T]}\right\}$$

Example 5.3 (log utility) If $u(w) = \ln(w)$, then the optimal attainable wealth is

$$W = vB_T/L$$

and the optimal objective value is

$$Eu(W) = \ln(v) - E[\ln(L/B_T)]$$

Example 5.4 (isoelastic utility) If $u(w) = \gamma^{-1}w^{\gamma}$, where $-\infty < \gamma < 1$ and $\gamma \neq 0$, then the optimal attainable wealth is

$$W = \frac{v(L/B_T)^{-1/(1-\gamma)}}{E[(L/B_T)^{-\gamma/(1-\gamma)}]}$$

and the optimal objective value is

$$Eu(W) = \frac{v^{\gamma}}{\gamma}\left\{E\left[(L/B_T)^{-\gamma/(1-\gamma)}\right]\right\}^{1-\gamma}$$

Example 5.5 (quadratic utility) This example builds on the results in section 2.4. If $u(w) = \beta w - w^2/2$ for parameter $\beta > 0$, then $I(i) = \beta - i$. Equation (5.9) becomes $W = \beta - \lambda L/B_T$. Solving equation (5.10) for λ and then substituting gives

$$W = \beta + \left[\frac{v - \beta E_Q[1/B_T]}{E_Q[L/B_T^2]}\right] L/B_T$$

for the optimal attainable wealth. Substituting this into the objective function eventually leads to

$$Eu(W) = \frac{\beta^2\left[E_Q[L/B_T^2] - \{E_Q[1/B_T]\}^2\right] - v^2 + 2\beta v E_Q[1/B_T]}{2E_Q[L/B_T^2]}$$

for the optimal objective value.

Example 5.1 and 5.2 (continued) The risk neutral probability measure and the state price vector are easily computed to be as follows:

ω	$Q(\omega)$	$L(\omega) = Q(\omega)/P(\omega)$
ω_1	$(1+5r)(2+8r)/12$	$(1+5r)(2+8r)/3$
ω_2	$(1+5r)(1-8r)/12$	$(1+5r)(1-8r)/3$
ω_3	$(3-5r)(1+4r)/12$	$(3-5r)(1+4r)/3$
ω_4	$(3-5r)(2-4r)/12$	$(3-5r)(2-4r)/3$

We first compute $E[L/B_2] = E_Q[(1+r)^{-2}] = (1+r)^{-2}$,

$$E[(L/B_2)\ln(L/B_2)] = (1+r)^{-2}E_Q[\ln(L)] - 2(1+r)^{-2}\ln(1+r)$$

and

$$\begin{aligned}E_Q[\ln(L)] = \frac{1}{12}\Big(&-12\ln(3) + 3(1+5r)\ln(1+5r) + 3(3-5r)\ln(3-5r)\\ &+ (1+5r)(2+8r)\ln(2+8r) + (1+5r)(1-8r)\ln(1-8r)\\ &+ (3-5r)(1+4r)\ln(1+4r) + (3-5r)(2-4r)\ln(2-4r)\Big)\end{aligned}$$

Hence the optimal attainable wealth is

$$W(\omega) = v(1+r)^2 + E_Q[\ln(L)] + \ln(3) + \begin{cases} -\ln(1+5r) - \ln(2+8r), & \omega = \omega_1 \\ -\ln(1+5r) - \ln(1-8r), & \omega = \omega_2 \\ -\ln(3-5r) - \ln(1+4r), & \omega = \omega_3 \\ -\ln(3-5r) - \ln(2-4r), & \omega = \omega_4 \end{cases}$$

Solving the system

$$(1+r)^2H_0(2) + 9H_1(2) = W(\omega_1)$$

$$(1+r)^2H_0(2) + 6H_1(2) = W(\omega_2)$$

yields

$$H_1(2) = -\frac{1}{3}\ln\left(\frac{2+8r}{1-8r}\right)$$

and

$$H_0(2) = v + \frac{E_Q[\ln(L)] + \ln(3)}{(1+r)^2} + \frac{2\ln(2+8r) - \ln(1+5r) - 3\ln(1-8r)}{(1+r)^2}$$

in states ω_1 and ω_2. Similarly, solving the system

$$(1+r)^2H_0(2) + 6H_1(2) = W(\omega_3)$$

$$(1+r)^2H_0(2) + 3H_1(2) = W(\omega_4)$$

yields

$$H_1(2) = \frac{1}{3}\ln\left(\frac{2-4r}{1+4r}\right)$$

and

$$H_0(2) = v + \frac{E_Q[\ln(L)] + \ln(3)}{(1+r)^2} + \frac{\ln(1+4r) - \ln(3-5r) - 2\ln(2-4r)}{(1+r)^2}$$

in states ω_3 and ω_4.

Next, solving the system

$$(1+r)H_0(1) + 8H_1(1) = V_1(\omega_1) = (1+r)H_0(2,\omega_1) + 8H_1(2,\omega_1)$$
$$(1+r)H_0(1) + 4H_1(1) = V_1(\omega_3) = (1+r)H_0(2,\omega_3) + 4H_1(2,\omega_3)$$

yields the values of $H_0(1)$ and $H_1(1)$ (see the calculations in section 5.1). Finally,

$$Eu(W) = 1 - (1+r)^{-2}\exp\{-v(1+r)^2 - E_Q[\ln(L)] + 2\ln(1+r)\}$$

is the optimal objective value.

Example 5.6 Suppose there is a single risky security that is governed by the binomial model with general parameters r, p, u, and d, and suppose $u(w) = \ln(w)$. In view of section 3.5,

$$L(\omega) = \frac{Q(\omega)}{P(\omega)} = \left(\frac{q}{p}\right)^n \left(\frac{1-q}{1-p}\right)^{T-n}$$

where $q = (1+r-d)/(u-d)$ and n is the number of "up" moves by the risky security corresponding to state ω. Recall that N_T, the number of up moves during the T periods, is a binomial random variable having parameters T and p. With log utility, it follows from example 5.3 that

$$W = v(1+r)^T \left(\frac{p}{q}\right)^{N_T} \left(\frac{1-p}{1-q}\right)^{T-N_T}$$

is the optimal attainable wealth. Moreover, since $EN_T = pT$, the optimal objective value is

$$\begin{aligned} Eu(W) &= \ln(v) + \ln(1+r)^T - E\ln(L) \\ &= \ln(v) + T\ln(1+r) - E\left[\ln\left(\frac{q}{p}\right)^{N_T}\right] - E\left[\ln\left(\frac{1-q}{1-p}\right)^{T-N_T}\right] \\ &= \ln(v) + T\ln(1+r) - pT\ln\left(\frac{q}{p}\right) - (1-p)T\ln\left(\frac{1-q}{1-p}\right) \end{aligned}$$

Now for arbitrary $n < T$, suppose $N_{T-1} = n$ and consider the optimal positions in the bank account and risky security that should be carried forward from time $T-1$. These are obtained by solving

$$(1+r)^T H_0(T) + S_{T-1}uH_1(T) = v(1+r)^T (p/q)^{n+1}(\{1-p\}/\{1-q\})^{T-n-1}$$
$$(1+r)^T H_0(T) + S_{T-1}dH_1(T) = v(1+r)^T (p/q)^{n}(\{1-p\}/\{1-q\})^{T-n}$$

thereby yielding

$$H_1(T) = \frac{v(1+r)^T (p/q)^n(\{1-p\}/\{1-q\})^{T-n-1}(p-q)}{S_{T-1}(u-d)q(1-q)}$$

and

$$H_0(T) = \frac{v(p/q)^n(\{1-p\}/\{1-q\})^{T-n-1}[u(1-p)q - d(1-q)p]}{(u-d)q(1-q)}$$

Since $V_{T-1} = (1+r)^{T-1}H_0(T) + S_{T-1}H_1(T)$, it follows with some algebra that

$$V_{T-1} = v(1+r)^{T-1}(p/q)^n(\{1-p\}/\{1-q\})^{T-n-1}$$

Now consider the fraction of money that is invested at time $T-1$ in the risky security. This is

$$\frac{S_{T-1}H_1(T)}{V_{T-1}} = \frac{(1+r)(p-q)}{(u-d)q(1-q)} \tag{5.11}$$

Notice this is independent of n and T. Moreover, notice that V_{T-1} has the same form as $V_T = W$, so an induction argument can be used to show that the optimal trading strategy has a very simple form: at each time and in each state simply invest the fraction (5.11) of one's wealth in the risky asset.

Exercise 5.5 For the model in example 5.1 with constant interest rate r equal to a general parameter $0 \leqslant r < 0.125$ and with log utility $u(w) = \ln(w)$, compute the optimal attainable wealth, the optimal objective value, and the optimal trading strategy using the risk neutral computational approach.

Exercise 5.6 In the specific case of example 5.1 with $r = 0$, compute the optimal attainable wealth, the optimal objective value, and the optimal trading strategy under the utility function

(a) $u(w) = -w^{-1}$
(b) $u(w) = \beta w - w^2/2$.

Exercise 5.7 Suppose there is a single risky security that is governed by the binomial model over T periods with constant interest rate $r \geqslant 0$ and general values for the parameters S_0, p, u, and d. Compute the optimal attainable wealth, the optimal objective value, and the optimal trading strategy under the exponential utility function $u(w) = -\exp(-w)$. In particular, show that if time t node entails n "ups" and $t-n$ "downs," then under the optimal strategy the corresponding value of the portfolio is

$$v(1+r)^t + \frac{[qt-n]\ln(q/p) + [(1-q)t + (n-t)]\ln((1-q)/(1-p))}{(1+r)^{T-t}}$$

Exercise 5.8 For the binomial model as in exercise 5.7 and with the quadratic utility function $u(w) = \beta w - w^2/2$, show that the optimal attainable wealth is

$$W = \beta + \frac{(1+r)^T v - \beta}{[q^2/p + (1-q)^2/(1-p)]^T}\left(\frac{q}{p}\right)^n\left[\frac{1-q}{1-p}\right]^{T-n}$$

where n is the number of "up" moves in the corresponding sample path. Moreover, show that the optimal objective value is

$$Eu(W) = \beta^2/2 - \frac{[(1+r)^T v - \beta]^2}{2[q^2/p + (1-q)^2/(1-p)]^T}$$

Hint:

$$\sum_{n=0}^{T}\binom{T}{n}a^n b^{T-n} = (a+b)^T$$

Exercise 5.9 For the binomial model as in exercises 5.7 and 5.8 and with the isoelastic utility function $u(w) = \gamma^{-1}w^\gamma$, show that the optimal attainable wealth is

$$W = \frac{(1+r)^T vL^{-1/(1-\gamma)}}{\left[p\left(\frac{q}{p}\right)^{-\gamma/(1-\gamma)} + (1-p)\left(\frac{1-q}{1-p}\right)^{-\gamma/(1-\gamma)}\right]^T}$$

and the optimal objective value is

$$Eu(W) = \frac{1}{\gamma}\left[(1+r)^T v\right]^\gamma\left[p\left(\frac{q}{p}\right)^{-\gamma/(1-\gamma)} + (1-p)\left(\frac{1-q}{1-p}\right)^{-\gamma/(1-\gamma)}\right]^{T(1-\gamma)}$$

Hint: see exercise 5.8.

5.3 Consumption–Investment and Dynamic Programming

A *consumption process* $C = \{C_t;\ t = 0, \dots, T\}$ is a non-negative, adapted stochastic process with C_t representing the amount of funds consumed by the investor at time t. A *consumption–investment plan* consists of a pair (C, H), where C is a consumption process and H is a trading strategy. A utility will be earned for the amount that is consumed in each period; naturally, the higher the consumption, the higher the utility. The investor seeks to choose the consumption–investment plan that maximizes the expected utility over the T periods. In particular, the investor faces a trade-off between consumption and investment, especially in the early periods. This section will explain how to solve this problem with dynamic programming.

Given the investor's initial wealth v, the consumption–investment plan (C, H) will be called *self-financing* if no money is added to or withdrawn from the portfolio between times 0 and T, other than the amounts that are consumed. As usual,

$$(5.12)\qquad V_t = H_0(t)B_t + \sum_{n=1}^{N} H_n(t)S_n(t), \quad t \geqslant 1$$

represents the value of the portfolio before any time t transactions. It will be assumed for $t \geqslant 1$ that V_t is also the value of the portfolio *before* any time t consumption. We let $V_0 = v$ denote the initial wealth. Thus to say (C, H) is self-financing means that

$$(5.13)\qquad V_t = C_t + H_0(t+1)B_t + \sum_{n=1}^{N} H_n(t+1)S_n(t), \quad t = 0, \dots, T-1$$

Given initial wealth v, the self-financing consumption–investment plan (C, H) is said to be *admissible* if $C_T \leqslant V_T$. Since C is a non-negative process, this implies $V_T \geqslant 0$.

The investor's consumption–investment problem is

$$(5.14)\qquad \begin{aligned} &\text{maximize} \quad E\left[\sum_{t=0}^{T}\alpha^t u(C_t)\right] \\ &\text{subject to:} \quad v = \text{initial wealth} \\ &\qquad\qquad (C, H) \text{ is admissible} \end{aligned}$$

Here $u : \mathbb{R} \to [-\infty, \infty]$ is a specified concave increasing utility function and α is a specified scalar parameter satisfying $0 < \alpha \leqslant 1$. Since the consumption process is required to be non-negative, without loss of generality it will be assumed that $u(w) = -\infty$ for all $w < 0$ (whereas, of course, $u(w) > -\infty$ for all $w > 0$).

To solve this problem with dynamic programming we will compute, working backwards in time in a recursive manner, the value function $u_t(w)$. This represents the maximum expected utility of consumption through time T, starting with wealth w and consumption at time t and given the time t history $\mathscr{F}_t$.

The value of u_T is easy to specify. Since the utility function u is increasing, the investor will want to consume all the wealth that is available in the final period. Thus $u_T = u$.

Starting at time $T-1$ with wealth w, the investor is faced with a problem that is essentially equivalent to a series of the single period problems of section 2.3:

$$\text{(5.15)} \quad \text{maximize} \quad u(C_{T-1}) + E[\alpha u_T(W)|\mathscr{F}_{T-1}]$$

$$\text{subject to: } w = C_{T-1} + H_0(T)B_{T-1} + \sum_{n=1}^{N} H_n(T)S_n(T-1)$$

$$W = H_0(T)B_T + \sum_{n=1}^{N} H_n(T)S_n(T)$$

$$H_n(T) \in \mathscr{F}_{T-1} \text{ for } n = 0, \ldots, N; \quad C_{T-1} \in \mathscr{F}_{T-1}$$

Notice the assumption that $u(w) = -\infty$ for all $w < 0$ will force the solution of (5.15) to satisfy $C_{T-1} \geqslant 0$ and $W \geqslant 0$.

Using the first constraint to solve for $H_0(T)$ and then substituting this in the second constraint, one eventually obtains

$$W = (w - C_{T-1})B_T/B_{T-1} + \sum_{n=1}^{N} H_n(T)[S_n(T) - B_T S_n(T-1)/B_{T-1}]$$

$$= (w - C_{T-1})B_T/B_{T-1} + B_T \sum_{n=1}^{N} H_n(T)\Delta S_n^*(T)$$

Hence (5.15) can be rewritten as

$$\max u(C_{T-1}) + \alpha E\left[u_T\left((w - C_{T-1})B_T/B_{T-1} + B_T \sum_{n=1}^{N} H_n(T)\Delta S_n^*(T)\right)|\mathscr{F}_{T-1}\right]$$

$$\text{subject to: } H_n(T) \in \mathscr{F}_{T-1} \text{ for } n = 1, \ldots, N \text{ and } C_{T-1} \in \mathscr{F}_{T-1}$$

We now set $u_{T-1}(w)$ equal to this optimal objective value.

In general, having computed $u_t(w)$, the value function $u_{t-1}(w)$ is computed from the *dynamic programming functional equation*

$$\text{(5.16)} \quad u_{t-1}(w) = \max\left\{u(C_{t-1}) + \alpha E\left(u_t\left((w - C_{t-1})B_t/B_{t-1} + B_t \sum_{n=1}^{N} H_n(t)\Delta S_n^*(t)\right)|\mathscr{F}_{t-1}\right)\right\}$$

where the maximum is over all $H_n(t) \in \mathscr{F}_{t-1}$ for $n = 1, \ldots, N$ and $C_{t-1} \in \mathscr{F}_{t-1}$. The value function $u_0(v)$ will then be the optimal objective value for the original problem (5.14) or (5.15), and the maximizing values of C_{t-1} and $H_n(t)$ will be part of the optimal consumption–investment plan. The final component, H_0, will come from the self-financing equation.

Example 5.7 Consider the security model of example 5.1 with interest rate the constant $r \geqslant 0$ and utility function $u(w) = u_2(w) = \ln(w)$. The dynamic programming functional equation for $t = 2$ and for states ω_1 and ω_2 is

$$\text{(5.17)} \quad u_1(w) = \max\{\ln(c) + \frac{\alpha}{2}\ln[(w-c)(1+r) + (1-8r)h] + \frac{\alpha}{2}\ln[(w-c)(1+r) - (2+8r)h]\}$$

Computing the partial derivatives of the argument with respect to c and h and then setting these equal to zero gives two equations; the solution is eventually found to be

$$c = \frac{w}{1+\alpha} \quad \text{and} \quad h = -\frac{1}{2}\frac{\alpha(1+r)(1+16r)w}{(1+\alpha)(2+8r)(1-8r)}$$

Note this means that if $S_1 = 8$ and the wealth at time 1 is w, then it is optimal to consume $w/(1+\alpha)$ right away and invest the balance by taking the (short) position h in the risky security and the (self-financing) position

$$H_0(2) = \frac{\alpha w/(1+\alpha) - 8h}{1+r}$$

in the bank account. Substituting these values of c and h back into the dynamic programming equation (5.17) eventually results in the time 1 value function for states ω_1 and ω_2:

$$u_1(w) = (1+\alpha)\ln(w) + f_8(\alpha, r)$$

where $f_8(\alpha, r)$ is a new function defined for convenience by

$$f_8(\alpha, r) = \alpha \ln\left(\frac{3\alpha(1+r)}{2(1+\alpha)(1-8r)}\right) + \frac{\alpha}{2}\ln\left(\frac{1-8r}{2+8r}\right) - \ln(1+\alpha)$$

Similarly, if $S_1 = 4$, then one computes the optimal time 1 consumption to be $c = w/(1+\alpha)$, the optimal position in the risky security:

$$h = \frac{1}{2}\frac{\alpha(1+r)(1-8r)w}{(1+\alpha)(2-4r)(1+4r)}$$

and the time 1 value function for states ω_3 and ω_4:

$$u_1(w) = (1+\alpha)\ln(w) + f_4(\alpha, r)$$

where $f_4(\alpha, r)$ is a new function defined for convenience by

$$f_4(\alpha, r) = \alpha \ln\left(\frac{3\alpha(1+r)}{2(1+\alpha)(1+4r)}\right) + \frac{\alpha}{2}\ln\left(\frac{1+4r}{2-4r}\right) - \ln(1+\alpha)$$

To summarize matters at this point,

$$u_1(w) = \begin{cases} (1+\alpha)\ln(w) + f_8(\alpha, r), & \omega = \omega_1, \omega_2 \\ (1+\alpha)\ln(w) + f_4(\alpha, r), & \omega = \omega_3, \omega_4 \end{cases}$$

and we are now in a position to use the dynamic programming equation (5.16) to recursively compute u_0:

$$u_0(w) = \max_{c,\, h} \left\{ \ln(c) + \frac{\alpha}{2}(1+\alpha)\ln[(w-c)(1+r)+(3-5r)h] + \frac{\alpha}{2}f_8 + \frac{\alpha}{2}(1+\alpha)\ln[(w-c)(1+r)-(1+5r)h] + \frac{\alpha}{2}f_4 \right\}$$

Computing the partial derivatives and so forth eventually leads to

$$c = \frac{w}{1+\alpha+\alpha^2} \quad \text{and} \quad h = \frac{\alpha(1+\alpha)(1+r)(1-5r)w}{(1+\alpha+\alpha^2)(3-5r)(1+5r)}$$

in which case

$$u_0(w) = (1+\alpha+\alpha^2)\ln(w) - \ln(1+\alpha+\alpha^2) + \alpha(1+\alpha)\ln\left(\frac{2\alpha(1+\alpha)(1+r)}{(1+\alpha+\alpha^2)}\right) - \frac{\alpha}{2}(1+\alpha)\ln[(1+5r)(3-5r)] + \frac{\alpha}{2}f_8(\alpha, r) + \frac{\alpha}{2}f_4(\alpha, r)$$

If you relax the assumption that $u(c) = -\infty$ for all $c \leqslant 0$, then the dynamic programming method can still be utilized to compute optimal consumption–investment plans, provided you explicitly worry about the constraint that the consumption process be non-negative. This may entail considerable extra work, as can be seen in the following example.

Example 5.8 Consider the security model of examples 5.1 and 5.7 with interest rate $r = 0$, $\alpha = 1$, and exponential utility function $u(c) = -\exp(-c)$. Since $u'(0) = 1$, in order to guard against negative consumption levels it is necessary to explicitly impose in the dynamic programming functional equation not only the constraint that the decision variable $c \geqslant 0$, but also constraints which guarantee that next period's wealth will be non-negative. For example, for states ω_1 and ω_2 and time $t = 1$ the dynamic programming equation should be

$$u_1(w) = \max_{\substack{c \geqslant 0 \\ w-c+h \geqslant 0 \\ w-c-2h \geqslant 0}} \left\{ -e^{-c} - \frac{1}{2}e^{-w+c-h} - \frac{1}{2}e^{-w+c+2h} \right\}$$

These three constraints define the feasible region, a triangular subset of $\mathbb{R}^2$ with vertices at $(c, h) = (0, -w)$, $(0, w/2)$, and $(w, 0)$.

We compute the maximizing values of the decision variables c and h by examining the partial derivatives. The partial derivative of the argument with respect to h is equal to zero if and only if $h = -\frac{1}{3}\ln 2$. The partial derivative with respect to c is equal to zero if and only if

$$c = \frac{1}{2}\{w + \ln 2 - \ln[e^{-h} + e^{2h}]\}$$

Substituting $h = -\frac{1}{3}\ln 2 = -0.231$, we see that the point $(c, h) = (0.0283 + w/2, -0.231)$ maximizes the argument over $\mathbb{R}^2$. This point will fall in the feasible region if and only if $w \geqslant 0.5186$, in which case, we see by substitution, $u_1(w) = -1.9442\, e^{-w/2}$.

If $w = 0.5186$, then the solution just discussed will make the constraint $w - c + h \geqslant 0$ tight. We therefore realize that the solution (c, h) must satisfy $c = w + h$ for all $w \leqslant 0.5186$. Substituting this equation, the objective function becomes $-e^{-w-h} - \frac{1}{2} - \frac{1}{2}e^{3h}$. The derivative of this equals zero if and only if $h = -w/4 - 0.1014$, in which case $c = 3w/4 - 0.1014$. But note that this gives $c \geqslant 0$ if and only if $w \geqslant 0.1352$. We therefore conclude that the optimal solution for all w satisfying $0.1352 \leqslant w \leqslant 0.5186$ is $(c, h) = (3w/4 - 0.1014, -w/4 - 0.1014)$, in which case, we see by substitution, $u_1(w) = -1.4756\, e^{-3w/4} - \frac{1}{2}$.

For $w = 0.1352$ this solution gives $c = 0$. We therefore realize that for all w satisfying $0 \leqslant w \leqslant 0.1352$ the optimal solution must be at the vertex of the triangle where $c = 0$ and $h = -w$, in which case $u_1(w) = -\frac{3}{2} - \frac{1}{2}e^{-3w}$.

Needless to say, the computation of $u_1(w)$ for states ω_3 and ω_4 will be equally difficult (see exercise 5.13). Moreover, with the messy nature of $u_1(w)$, the computation of $u_0(w)$ will be even more difficult. Without taking this example further, it should be clear that the computational difficulties can become formidable when you relax the assumption that $u'(0) = \infty$.

Exercise 5.10 For the security model of example 5.1 with constant interest rate $r = 0$, utility parameter $\alpha = 1$, and isoelastic utility function $u(c) = -1/c$, use dynamic programming to compute the optimal consumption process and the optimal trading strategy.

Exercise 5.11 For a single security governed by the binomial model with parameters p, u, d, r, and T, with utility parameter $\alpha \leqslant 1$, and with log utility $u(c) = \ln(c)$, use dynamic programming to show that the optimal amount to consume at time t is

$$C_t = \frac{W_t}{1 + \alpha + \ldots + \alpha^{T-t}}$$

where W_t is the wealth available at that same time. Moreover, show that at every time t the optimal fraction of the invested funds which are put in the risky security is

$$\frac{S_t H_1(t+1)}{W_t - C_t} = \frac{(1+r)[pu + (1-p)d - (1+r)]}{(1+r-d)(u-1-r)}$$

Finally, show that the time t value function is of the form $u_t(w) = (1+\alpha+\ldots+\alpha^{T-t})\ln(w)$ plus a constant.

Exercise 5.12 For a single security governed by the binomial model with parameters p, u, d, r, and T, with utility parameter $\alpha \leqslant 1$, and with isoelastic utility $u(c) = c^\gamma/\gamma$, use dynamic programming and an induction argument to show that the optimal amount to consume at time t is $C_t = k_t W_t$, where k_t is a positive constant that depends on t but is independent of the state ω. Show that at every time t the optimal fraction of the invested funds which are put in the risky security is

$$\frac{S_t H_1(t+1)}{W_t - C_t} = \frac{(1+r)\left[(1-p)^{1/(\gamma-1)}(1+r-d)^{1/(\gamma-1)} - p^{1/(\gamma-1)}(u-1-r)^{1/(\gamma-1)}\right]}{-(1-p)^{1/(\gamma-1)}(1+r-d)^{\gamma/(\gamma-1)} + p^{1/(\gamma-1)}(u-1-r)^{\gamma/(\gamma-1)}}$$

Finally, show that the time t value function is of the form $u_t(w) = \gamma_t w^\gamma$, where γ_t is a constant that depends on t but is independent of the state ω.

Exercise 5.13 For the situation in example 5.8, compute $u_1(w)$ and the corresponding maximizing values of c and h for states ω_3 and ω_4 and for all $w \geqslant 0$.

5.4 Consumption–Investment and Martingale Methods

As an alternative to dynamic programming, the risk neutral probability measure can be exploited to provide an efficient method for solving the consumption–investment problem (5.14), an approach that is a natural generalization of the one taken for single period models. Throughout this section it will be assumed that there exists a unique martingale measure Q, so the model is complete. A consumption process C will be called *attainable* if there exists a trading strategy H such that (C, H) is an admissible consumption–investment plan satisfying $C_T = V_T$ (implicit is a specified initial wealth V). In this case one says that H *replicates* or *generates* C.

The first step in the risk neutral computational approach is to characterize the set of all attainable consumption processes. The second step is to find the element of this set that maximizes expected utility, that is, the objective function in (5.14). Finally, one derives the self-financing trading strategy H that generates this optimal C.

We begin by observing the following:

(5.18) Given an initial wealth $v \geqslant 0$, a consumption process C, and a self-financing trading strategy H, one has

$$V_t/B_t = v + G_t^* - \sum_{u=0}^{t-1} C_u/B_u, \qquad t = 1, \ldots, T$$

One can see this by an induction argument. The reader can use (5.13) to verify this is true for $t = 1$. For the induction step, suppose this equation holds for $t = s$. Now (5.13) implies

$$H_0(s+1) = V_s/B_s - C_s/B_s - \sum_{n=1}^{N} H_n(s+1)S_n^*(s)$$

so substituting this in (5.12) with $t = s+1$ after dividing by B_{s+1} yields

$$V_{s+1}/B_{s+1} = V_s/B_s + \sum_{n=1}^{N} H_n(s+1)\Delta S_n^*(s+1) - C_s/B_s$$

The induction assumption that (5.18) holds for $t = s$ therefore finishes the argument that (5.18) holds for $t = s+1$.

Relationship (5.18) will enable us to characterize the attainable consumption processes. If we define $M_t = v + G_t^*$, then M is a martingale under the risk neutral probability measure Q satisfying $M_0 = v$. Hence if (5.18) applies, then $M_t = V_t/B_t + C_0/B_0 + \ldots + C_{t-1}/B_{t-1}$ and

$$v = E_Q[V_t/B_t + C_0/B_0 + \ldots + C_{t-1}/B_{t-1}], \qquad t = 1, \ldots, T$$

If, in addition, $C_T = V_T$, then

(5.19) $$v = E_Q[C_0/B_0 + \ldots + C_T/B_T]$$

In other words, equation (5.19) is a necessary condition for the consumption process C to be attainable. It turns out that (5.19) is also a sufficient condition, that is,

(5.20) Given the initial wealth $v \geqslant 0$, the consumption process C is attainable if and only if (5.19) holds. If V_0 is the value of the portfolio which replicates C, then $V_0 \geqslant 0$.

To see why equation (5.19) is sufficient, observe that $X \equiv B_T[C_0/B_0 + \ldots + C_T/B_T]$ is an attainable contingent claim. Actually, X should be viewed as the composition of $T+1$ attainable contingent claims, where the t^{th} claim is the receipt of C_t dollars at time t, which are then deposited and held in the bank account until time T. Thus there are T self-financing trading strategies $H^1, \ldots, H^T$ that replicate the T contingent claims $C_1, \ldots, C_T$, respectively. Taking $H \equiv H^1 + \ldots + H^T$, it follows that H is a self-financing trading strategy such that (C, H) is an admissible consumption–investment plan with $C_T = V_T$.

For the second part of (5.20), we see from (5.18) that

$$V_t/B_t + \sum_{u=0}^{t-1} C_u/B_u = E_Q\left[V_T/B_T + \sum_{u=0}^{T-1} C_u/B_u \mid \mathscr{F}_t\right]$$

is a martingale under Q with time T value equal to

$$\sum_{u=0}^{T} C_u/B_u$$

Thus

$$V_t/B_t = E_Q\left[\sum_{u=t}^{T} C_u/B_u \mid \mathscr{F}_t\right]$$

must be non-negative for all t.

Assuming the utility function satisfies $u(c) = -\infty$ for all $c < 0$, then by (5.20) the optimal consumption–investment problem (5.14) is equivalent to the following:

$$\text{(5.21)} \qquad \begin{aligned} &\text{maximize } E\left[\sum_{t=0}^{T} \alpha^t u(C_t)\right] \\ &\text{subject to } E_Q[C_0/B_0 + \ldots + C_T/B_T] = v \\ &\qquad C \text{ is an adapted process} \end{aligned}$$

The assumption about the utility function will guarantee that the optimal solution is a non-negative stochastic process. In view of (5.20), the optimal solution will be an attainable consumption process whose objective value is greater than or equal to that for every other attainable consumption process. Hence with a solution C of (5.21), all that remains to obtain the solution of (5.14) is to derive the trading strategy that replicates C.

Problem (5.21) can be solved with a Lagrange multiplier in a fashion similar to the optimal portfolio problem of section 5.2. However, the situation here is a little more tricky because the decision variable is an adapted process, not just a random variable. The following result will play a crucial role.

$$\text{(5.22)} \qquad E_Q\left[\sum_{t=0}^{T} C_t/B_t\right] = E\left[\sum_{t=0}^{T} C_t N_t\right]$$

where N is the adapted stochastic process defined for all t by $N_t = E[L|\mathscr{F}_t]/B_t$.

Since $L = Q/P$, this result follows from a simple calculation:

$$\begin{aligned} E_Q\left[\sum_{t=0}^{T} C_t/B_t\right] &= E\left[L\sum_{t=0}^{T} C_t/B_t\right] \\ &= E\left[\sum_{t=0}^{T} E[C_t L/B_t|\mathscr{F}_t]\right] = E\left[\sum_{t=0}^{T} C_t N_t\right] \end{aligned}$$

As a consequence, problem (5.21) can be rewritten as follows:

$$\text{(5.23)} \qquad \begin{aligned} &\text{maximize } \quad E\left[\sum_{t=0}^{T} \alpha^t u(C_t)\right] \\ &\text{subject to } \quad E\left[\sum_{t=0}^{T} C_t N_t\right] = v \\ &\qquad C \text{ is an adapted process.} \end{aligned}$$

Introducing a Lagrange multiplier λ, we now want to solve:

$$\text{(5.24)} \qquad \text{maximize } E\left[\sum_{t=0}^{T} \alpha^t u(C_t) - \lambda \sum_{t=0}^{T} C_t N_t\right]$$

With suitable assumptions about the utility function u to ensure the optimal solution C will feature strictly positive consumption values (e.g., it suffices to require that the marginal utility $u'(c)$ converges to ∞ as c approaches 0 from above and that $u'(c)$ converges to 0 as c increases to ∞), the following first order necessary condition must be satisfied:

$$\text{(5.25)} \qquad \alpha^t u'(C_t) = \lambda N_t, \quad \text{all} \quad \omega \in \Omega, \quad t = 0, \ldots, T$$

Equivalently, if $I(\cdot)$ is the inverse of the marginal utility function $u'(\cdot)$, then we must have

$$\text{(5.26)} \qquad C_t = I(\lambda N_t/\alpha^t), \quad \text{all} \quad \omega \in \Omega, \quad t = 0, \ldots, T$$

All that remains is to establish the correct value of the Lagrange multiplier λ; this is simply the value such that the constraint in (5.23) is satisfied when (5.26) is substituted, that is, the correct value of λ is the unique solution of

$$\text{(5.27)} \qquad E\left[\sum_{t=0}^{T} N_t I(\lambda N_t/\alpha^t)\right] = v$$

Example 5.9 With $u(c) = \ln(c)$ we have $u'(c) = c^{-1}$ and $I(i) = i^{-1}$, in which case $I(\lambda N_t/\alpha^t) = \alpha^t/(\lambda N_t)$. Equation (5.27) becomes

$$v = E\left[\sum_{t=0}^{T} N_t \alpha^t/(\lambda N_t)\right] = \frac{1}{\lambda}\sum_{t=0}^{T} \alpha^t$$

Thus

$$\lambda = \begin{cases} \frac{T+1}{v}, & \alpha = 1 \\ \frac{1-\alpha^{T+1}}{v(1-\alpha)}, & \alpha < 1 \end{cases}$$

It follows from (5.26) that the optimal solution is

$$C_t = \frac{\alpha^t}{\lambda N_t} = \begin{cases} \frac{v}{(T+1)N_t}, & \alpha = 1; \quad t = 0, \ldots, T \\ \frac{\alpha^t v(1-\alpha)}{(1-\alpha^{T+1})N_t} & \alpha < 1; \quad t = 0, \ldots, T \end{cases}$$

Thus, for instance, when $\alpha = 1$ the optimal objective value is $(T+1) \ln\{v/(T+1)\} - E\ln(N_0) - \ldots - E\ln(N_T)$.

Examples 5.1 and 5.9 (continued) For the security model of example 5.1 with a constant interest rate $r \geqslant 0$, the stochastic process N is as follows:

ω	$P(\omega)$	$L(\omega)$	N_0	N_1	N_2
ω_1	1/4	$\frac{(1+5r)(2+8r)}{3}$	1	$\frac{1+5r}{2(1+r)}$	$\frac{L}{(1+r)^2}$
ω_2	1/4	$\frac{(1+5r)(1-8r)}{3}$	1	$\frac{1+5r}{2(1+r)}$	$\frac{L}{(1+r)^2}$
ω_3	1/4	$\frac{(3-5r)(1+4r)}{3}$	1	$\frac{(3-5r)}{2(1+r)}$	$\frac{L}{(1+r)^2}$
ω_4	1/4	$\frac{(3-5r)(2-4r)}{3}$	1	$\frac{3-5r}{2(1+r)}$	$\frac{L}{(1+r)^2}$

The optimal consumption process is $C_t = \alpha^t v/[(1+\alpha+\alpha^2)N_t]$, so, in particular,

$$C_2 = V_2 = \frac{\alpha^2 v(1+r)^2}{(1+\alpha+\alpha^2)L}$$

Now by considering how to replicate the contingent claim V_2, we must have, in states ω_1 and ω_2, respectively,

$$(1+r)^2 H_0(2) + 9H_1(2) = \frac{\alpha^2 v(1+r)^2}{(1+\alpha+\alpha^2)L(\omega_1)}$$

and

$$(1+r)^2 H_0(2) + 6H_1(2) = \frac{\alpha^2 v(1+r)^2}{(1+\alpha+\alpha^2)L(\omega_2)}$$

Solving for $H_0(2)$ and $H_1(2)$ yields

$$H_1(2) = -\frac{\alpha^2 v(1+r)^2(1+16r)}{(1+\alpha+\alpha^2)(1+5r)(2+8r)(1-8r)}$$

and

$$H_0(2) = \frac{12\alpha^2 v(1+10r)}{(1+\alpha+\alpha^2)(1+5r)(2+8r)(1-8r)}$$

Hence the value of the portfolio just after time 1 consumption is

$$(1+r)H_0(2) + 8H_1(2) = \frac{4\alpha^2 v(1+r)(1+4r)}{(1+\alpha+\alpha^2)(1+5r)(2+8r)}$$

Adding C_1 to this gives

$$V_1 = \frac{2\alpha v(1+r)(1+\alpha)}{(1+\alpha+\alpha^2)(1+5r)}$$

for the value of the portfolio just before time 1 consumption in states ω_1 and ω_2.

In a similar fashion we compute

$$H_1(2) = \frac{\alpha^2 v(1+r)^2(1-8r)}{(1+\alpha+\alpha^2)(3-5r)(1+4r)(2-4r)}$$

$$H_0(2) = \frac{36\alpha^2 vr}{(1+\alpha+\alpha^2)(3-5r)(1+4r)(2-4r)}$$

and

$$V_1 = \frac{2\alpha v(1+r)(1+\alpha)}{(1+\alpha+\alpha^2)(3-5r)}$$

in states ω_3 and ω_4.

Finally, we compute the trading strategy that will replicate V_1, giving

$$H_1(1) = \frac{\alpha(1+\alpha)(1+r)(1-5r)v}{(1+\alpha+\alpha^2)(1+5r)(3-5r)}$$

and

$$H_0(1) = \frac{2\alpha(1+\alpha)(-1+15r)v}{(1+\alpha+\alpha^2)(1+5r)(3-5r)}$$

Notice that $C_0 + H_0(1) + 5H_1(1) = v$, as required.

Exercise 5.14 Use the martingale approach to show that with the isoelastic utility function $u(c) = c^\gamma/\gamma, \gamma < 1, \gamma \neq 0$, the optimal consumption process is given by

$$C_t = \frac{v}{\Delta}\alpha^{t/(1-\gamma)} N_t^{1/(\gamma-1)}$$

and the optimal objective value is given by $v^\gamma \Delta^{1-\gamma}/\gamma$, where

$$\Delta = \sum_{t=0}^{T} \alpha^{t/(1-\gamma)} E\left[N_t^{\gamma/(\gamma-1)}\right]$$

Exercise 5.15 For the security model of example 5.1 with $\alpha \leqslant 1$, initial wealth v, constant interest rate $r \geqslant 0$, and isoelastic utility function $u(c) = c^\gamma/\gamma$, use the risk neutral computational approach to write algebraic formulas for the optimal consumption process C. Compute numerical values for this process as well as the optimal trading strategy in the special case $\gamma = -1$, $\alpha = 1$, and $r = 0$.

5.5 Maximum Utility From Consumption and Terminal Wealth

This section investigates the problem where utility is gained from both the consumption each period as well as the amount of money that is dedicated at time T for subsequent use. This is a modest generalization of the ordinary

consumption–investment problem, where now only a portion C_T of the time T wealth V_T is consumed, leaving $V_T - C_T$ for future investment. This new situation is equivalent to the ordinary consumption–investment problem, except that the utility function for period T consumption is allowed to be different from the utility function for the other periods. Not surprisingly, it is easy to solve this generalized consumption–investment problem, using slight generalizations of either the dynamic programming or the risk neutral computational approaches.

Throughout this section it will be assumed that there exists a unique risk neutral probability measure Q, so the model is complete. Let $\mathscr{A}_v$ denote the set of all the admissible consumption–investment plans with initial wealth $V_0 = v$, so each $(C, H) \in \mathscr{A}_v$ is self-financing with C a non-negative, adapted process satisfying $C_T \leqslant V_T$.

For the generalized consumption–investment problem, two concave, increasing utility functions are specified: u_c measures the utility of consumption each period and satisfies $u_c(c) = -\infty$ for all $c < 0$ (for computational convenience we shall usually assume, in addition, $u_c'(c) \to \infty$ as $c \searrow 0$), while u_p measures the utility of the funds that are reserved at time T for the future, as in an ordinary optimal portfolio problem. The problem is to choose the $(C, H) \in \mathscr{A}_v$ which maximizes the expected total utility, which is

$$E\left[\sum_{t=0}^{T} \alpha^t u_c(C_t) + \alpha^T u_p(V_T - C_T)\right]$$

We are also interested in the corresponding *value function*, which keeps track of the optimal objective value as a function of the initial wealth v. This is denoted $J(v)$ and is given by

$$\textbf{(5.28)} \qquad J(v) = \max_{(C,H) \in \mathscr{A}_v} E\left[\sum_{t=0}^{T} \alpha^t u_c(C_t) + \alpha^T u_p(V_T - C_T)\right]$$

With the dynamic programming approach we use (5.16), exactly the same recursive functional equation as with the ordinary consumption–investment problem, for $t = 1, \ldots, T$, namely

$$u_{t-1}(w) = \max_{C,H} \left\{ u_c(C) + \alpha E\Big(u_t\Big((w - C)B_t/B_{t-1} + B_t \sum_{n=1}^{N} H_n \Delta S_n^*(t)\Big) \Big| \mathscr{F}_{t-1}\Big)\right\}$$

where the maximum is taken over all $H_n \in \mathscr{F}_{t-1}$ for $n = 1, \ldots, N$ and over all $C \in \mathscr{F}_{t-1}$. As usual, the function $u_t(w)$ is to be interpreted as the maximum expected discounted utility beginning at time t with wealth w, and so $u_0(w) = J(w)$. Moreover, after recursively computing the u_t functions, one obtains the optimal consumption–investment plan by taking the maximizing values of H and C, beginning with $t = 1$.

A crucial difference between the dynamic programming approach for the ordinary consumption–investment problem and the one here is the specification of the terminal utility function u_T. In the former case one simply takes $u_t(w) = u_c(w)$, because with wealth w at time T the obvious thing to do is to consume everything. But now in this situation one needs to optimally divide the wealth between immediate consumption and terminal investment, which means one should take

$$\textbf{(5.29)} \qquad u_T(w) = \max_{0 \leqslant c \leqslant w} \{u_c(c) + u_p(w - c)\}$$

Hence the dynamic programming approach for a problem featuring both consumption and terminal wealth is essentially the same as for an ordinary consumption–investment problem, except that the utility associated with time T consumption should be as in (5.29) instead of u_c.

Example 5.10 Suppose $u_c(c) = \ln(c)$ and $u_p(w) = w^\gamma/\gamma$ with $\gamma < 1$ and $\gamma \neq 0$, so $u_T(w) = \max\{\ln(c) + (w - c)^\gamma/\gamma\}$. A little calculus shows the maximizing c is a root of the equation $c = (w - c)^{1-\gamma}$. For instance, if $\gamma = -1$, then

$$c = w + 1/2 - \sqrt{w + 1/4} = [\sqrt{w + 1/4} - 1/2]^2$$

in which case

$$\begin{aligned} u_T(w) &= 2\ln(\sqrt{w + 1/4} - 1/2) - \frac{1}{\sqrt{w + 1/4} - 1/2} \\ &= 2u_c(g(w)) + u_p(g(w)) \end{aligned}$$

where g is the concave, increasing function $g(w) = \sqrt{w + 1/4} - 1/2$ with domain $[0, \infty)$.

Turning to the risk neutral computational approach, in view of (5.19) and (5.20) it should be clear that the following holds:

(5.30) Given the initial wealth $v \geqslant 0$ and the admissible consumption–investment plan (C, H), one has

$$v = E_Q[C_0/B_0 + \ldots + C_{T-1}/B_{T-1} + V_T/B_T]$$

Conversely, if this equation holds and if C is a consumption process with $C_T \leqslant V_T$, then there exists a trading strategy H such that (C, H) is an admissible consumption–investment plan with $v \geqslant 0$.

It follows that our optimization problem can be formulated as follows:

$$\begin{aligned} \text{maximize} \quad & E\left[\sum_{t=0}^{T} \alpha^t u_c(C_t) + \alpha^T u_p(V_T - C_T)\right] \\ \text{subject to} \quad & v = E_Q[C_0/B_0 + \ldots + C_{T-1}/B_{T-1} + V_T/B_T] \\ & V_T \geqslant C_T, \ V_T \in \mathscr{F}_T \\ & C \text{ is an adapted process} \end{aligned}$$

Our assumption that $u_c'(c) \to \infty$ as $c \searrow 0$ will guarantee that the maximizing consumption process will satisfy $C_t > 0$ for all t. We will assume in addition that $u_p'(w) \to \infty$ as $w \searrow 0$, so that the constraint $V_T > C_T$ will automatically be satisfied by the optimal solution. In other words, with these assumptions about the utility functions it suffices to work with the following formulation of the optimization problem:

$$\text{(5.31)} \qquad \begin{aligned} &\text{maximize } E\left[\sum_{t=0}^{T} \alpha^t u_c(C_t) + \alpha^T u_p(V_T - C_T)\right] \\ &\text{subject to } v = E_Q[C_0/B_0 + \ldots + C_{T-1}/B_{T-1} + V_T/B_T] \\ &\qquad C \text{ is an adapted process} \\ &\qquad V_T \in \mathscr{F}_T \end{aligned}$$

The optimal solution will immediately give a consumption process with $C_t > 0$ for all t and with $C_T < V_T$. It follows from (5.30) that there exists a trading strategy H making (C, H) an admissible consumption–investment plan which must be the solution of the original optimization problem.

Problem (5.31) is solved with the same approach as was used with problem (5.21). We introduce the Lagrange multiplier λ and the adapted process N defined by $N_t = E[L|\mathscr{F}_t]/B_t$, thereby allowing us to rewrite (5.31) as

$$\text{maximize} \quad E\left[\sum_{t=0}^{T} \alpha^t u_c(C_t) + \alpha^T u_p(V_T - C_T) - \lambda \sum_{t=0}^{T-1} C_t N_t - \lambda V_T N_T\right]$$

where the maximum is over stochastic processes C and random variables V_T. Differentiating with respect to each C_t and then V_T, we see that the following first order necessary conditions must be satisfied:

$$\begin{aligned} \alpha^t u_c'(C_t) &= \lambda N_t, \qquad t = 0, \ldots, T-1 \\ \alpha^T u_c'(C_T) &= \alpha^T u_p'(V_T - C_T), \\ \alpha^T u_p'(V_T - C_T) &= \lambda N_T \end{aligned}$$

Introducing $I_c(\cdot)$, the inverse of the marginal utility function $u_c'(\cdot)$, as well as $I_p(\cdot)$, the inverse of the marginal utility function $u_p'(\cdot)$, it follows that the optimal solution must satisfy

$$\text{(5.32)} \qquad \begin{aligned} C_t &= I_c(\lambda N_t/\alpha^t), \qquad t = 0, \ldots, T, \\ V_T &= I_c(\lambda N_T/\alpha^T) + I_p(\lambda N_T/\alpha^T) \end{aligned}$$

for some positive value of the scalar λ. The correct value of λ is the one which satisfies the constraint in (5.31) upon substitution of (5.32), namely,

$$\text{(5.33)} \qquad E\left[\sum_{t=0}^{T} I_c(\lambda N_t/\alpha^t) N_t + I_p(\lambda N_T/\alpha^T) N_T\right] = v$$

With this value of λ the optimal objective value is thus

$$J(v) = E\left[\sum_{t=0}^{T} \alpha^t u_c(I_c(\lambda N_t/\alpha^t)) + \alpha^T u_p(I_p(\lambda N_T/\alpha^T))\right]$$

Example 5.10 (continued) Suppose $u_c(c) = \ln(c)$ and $u_p(w) = -1/w$, so $I_c(i) = 1/i$ and $I_p(i) = 1/\sqrt{i}$. Hence (5.33) becomes

$$\frac{1}{\lambda} \sum_{t=0}^{T} \alpha^t + \frac{1}{\sqrt{\lambda}} E\left[\sqrt{N_T \alpha^T}\right] = v$$

and so the correct value of λ can be obtained by solving a simple quadratic equation.

We can sharpen these results if the utility functions are sufficiently differentiable. Define $f : (0, \infty) \to (0, \infty)$ by

$$f(\lambda) = E\left[\sum_{t=0}^{T} I_c(\lambda N_t/\alpha^t) N_t + I_p(\lambda N_T/\alpha^T) N_T\right]$$

so equation (5.33) is the same thing as $f(\lambda) = v$. With minor assumptions about the utility functions, the function f will be continuous and strictly decreasing, in which case it will have an inverse function, which will be denoted g. Thus $f(\lambda) = v$ if and only if $\lambda = g(v)$. In particular, the expressions in (5.32) for the optimal solution become

$$\begin{aligned} C_t &= I_c(g(v) N_t/\alpha^t), \qquad t = 0, \ldots, T \\ V_T &= I_c(g(v) N_T/\alpha^T) + I_p(g(v) N_T/\alpha^T) \end{aligned}$$

and the optimal objective value $J(v) = K(g(v))$, where we have introduced the function

$$K(\lambda) \equiv E\left[\sum_{t=0}^{T} \alpha^t u_c(I_c(\lambda N_t/\alpha^t)) + \alpha^T u_p(I_p(\lambda N_T/\alpha^T))\right]$$

But

$$f'(\lambda) = E\left[\sum_{t=0}^{T} I_c'(\lambda N_t/\alpha^t) N_t^2/\alpha^t + I_p'(\lambda N_T/\alpha^T) N_T^2/\alpha^T\right]$$

so

$$\begin{aligned} K'(\lambda) = E\Bigg[&\sum_{t=0}^{T} \alpha^t u_c'(I_c(\lambda N_t/\alpha^t)) \frac{d}{d\lambda} I_c(\lambda N_t/\alpha^t) \\ &+ \alpha^T u_p'(I_p(\lambda N_T/\alpha^T)) \frac{d}{d\lambda} I_p(\lambda N_T/\alpha^T)\Bigg] \end{aligned}$$

$$= E\left[\sum_{t=0}^{T} \alpha^t(\lambda N_t/\alpha^t)\frac{d}{d\lambda}I_c(\lambda N_t/\alpha^t) + \alpha^T(\lambda N_T/\alpha^T)\frac{d}{d\lambda}I_p(\lambda N_T/\alpha^T)\right]$$

$$= E\left[\sum_{t=0}^{T} \alpha^t(\lambda N_t/\alpha^t)^2 I_c'(\lambda N_t/\alpha^t) + \alpha^T(\lambda N_T/\alpha^T)^2 I_p'(\lambda N_T/\alpha^T)\right]$$

$$= \lambda f'(\lambda)$$

It follows that

$$J'(v) = g(v)$$

because by standard rules of differential calculus

$$J'(v) = K'(g(v))g'(v) = g(v)f'(g(v))g'(v) = g(v)$$

Notice that the g function will be positive and strictly decreasing, so we conclude from these calculations that the value function J is strictly increasing and strictly concave.

Exercise 5.16 For a simple single period model with one security with $\Omega = \{\omega_1, \omega_2\}$, $S_0 = 4$, $S_1(\omega_1) = 5$, $S_1(\omega_2) = 3$, $\alpha = 1$ and constant interest rate $r = 0$, suppose the utility functions are as in example 5.10 with $\gamma = -1$. Moreover, suppose $P(\omega_1) = 2/3$. Compute the optimal consumption process and trading strategy for the generalized consumption–investment problem of this section using:

(a) the dynamic programming approach,
(b) the risk neutral computational approach.

Exercise 5.17 Show that the $g(v)$ functions corresponding to the isoelastic utility function $u(w) = \gamma^{-1}w^{\gamma}$ for general values of the parameter γ, where $-\infty < \gamma < 1$ and $\gamma \neq 0$, are given by

$$g_c(v) = v^{\gamma-1}\left(E\left[\sum_{t=0}^{T} N_t^{\gamma/(\gamma-1)}\alpha^{t/(1-\gamma)}\right]\right)^{1-\gamma}$$

and

$$g_p(v) = v^{\gamma-1}\left(E\left[L^{\gamma/(\gamma-1)}B_T^{\gamma/(1-\gamma)}\right]\right)^{1-\gamma}$$

for the ordinary consumption–investment and the ordinary optimal portfolio problems, respectively.

Exercise 5.18 Show that if $u_p(w) = \ln(w)$, then $g(v) = 1/v$ for an ordinary optimal portfolio problem with $u_c = 0$.

5.6 Optimal Portfolios With Constraints

The ideas in this section are, for the most part, straightforward generalizations of the single period concepts in section 2.5. A non-empty, closed, convex subset $\mathbb{K}$ of $\mathbb{R}^N$ is specified, and the trading strategy, when expressed as the N-vector representing the fractions of the wealth held in the risky securities, is required to be an element of $\mathbb{K}$ at each time t. Examples of $\mathbb{K}$ are given in section 2.5. This constrained, multiperiod problem can be solved with either dynamic programming or a risk neutral computational approach, as will now be discussed.

To be precise, *trading strategies* will be of the form $F = (F_1, \ldots, F_N)$, where each $F_n = \{F_n(t); t = 1, \ldots, T\}$ is a predictable stochastic process with $F_n(t) = H_n(t)S_n(t-1)/V_{t-1}$ representing the fraction of the time $t-1$ wealth that is invested at that time in security n and held until time t. Assuming the trading strategy is self-financing, it follows that $1 - F_1(t) - \ldots - F_N(t)$ is the fraction of time $t-1$ wealth that is invested in the bank account. In general, the value $F_n(t)$ can be less than zero or greater than one. However, with the specification of the constraint set $\mathbb{K}$, for a trading strategy to be *admissible* it is necessary that $F(t) \in \mathbb{K}$ for $t = 1, \ldots, T$. Let $\mathscr{A}$ denote the set of all such admissible trading strategies.

According to section 3.2, the time T value V_T of the portfolio can be expressed as

$$V_T = v\prod_{t=1}^{T}\left[1 + r_t + \sum_{n=1}^{N} F_n(t)\{\Delta R_n(t) - r_t\}\right]$$

where v is the initial wealth, $r_t = (B_t - B_{t-1})/B_{t-1}$ is the interest rate associated with the interval $[t-1, t]$, and $\Delta R_n(t) = R_n(t) - R_n(t-1)$ is the change in the return process associated with security n. With the specification of the utility function u and the initial wealth v, the constrained optimal portfolio problem is:

$$\text{(5.34)} \qquad \begin{aligned} &\text{maximize} && Eu(V_T) \\ &\text{subject to} && F \in \mathscr{A}, \quad V_0 = v \end{aligned}$$

The corresponding *value function* is

$$J(v) \equiv \sup_{F \in \mathscr{A}} Eu(V_T)$$

This problem can be solved with dynamic programming by using recursively the dynamic programming functional equation:

$$u_{t-1}(w) = \max_{F \in \mathbb{K}} E\left[u_t\left(w\left\{1 + r_t + \sum_{n=1}^{N} F\{\Delta R_n(t) - r_t\}\right\}\right)\middle|\mathscr{F}_{t-1}\right]$$

for $t = 1, 2, \ldots, T$ together with the boundary condition $u_T(w) = u(w)$. Thus $u_t(w)$ represents the maximum expected utility of terminal wealth when you start with initial wealth w at time t and the information available is $\mathscr{F}_t$. In particular, $J(v) = u_0(v)$. The dynamic programming approach is illustrated in the following example.

Example 5.11 Consider the two-period, single security model in example 5.1 with the interest rate $r_t = 0$. Some relevant data are:

ω	$P(\omega)$	$\Delta R_1(\omega)$	$\Delta R_2(\omega)$
ω_1	1/4	3/5	1/8
ω_2	1/4	3/5	−2/8
ω_3	1/4	−1/5	2/4
ω_4	1/4	−1/5	−1/4

Suppose the utility function $u(w) = \ln(w)$ and short selling of the risky security is prohibited, so $\mathbb{K} = [0, \infty)$. The dynamic programming equation for $t = 2$ and $\omega = \omega_1, \omega_2$ is

$$u_1(w) = \max_{F \geqslant 0} E[\ln(w\{1 + F\Delta R_2\}|S_1 = 8]$$

$$= \max_{F \geqslant 0} \left\{\frac{1}{2}\ln[w(1 + F/8)] + \frac{1}{2}\ln[w(1 - F/4)]\right\}$$

The argument is a concave function with respect to F, and its derivative at $F = 0$ is negative, so the argument is maximized on $[0, \infty)$ at $F = 0$. Substituting this value gives $u_1(w) = \ln(w)$ for ω_1 and ω_2.

The dynamic programming equation for $t = 2$ and $\omega = \omega_3, \omega_4$ is

$$u_1(w) = \max_{F \geqslant 0} \left\{\frac{1}{2}\ln[w(1 + F/2)] + \frac{1}{2}\ln[w(1 - F/4)]\right.$$

The argument here is concave with respect to F, but the derivative at $F = 0$ is positive, so the argument is maximized on $[0, \infty)$ where the derivative equals 0. This is easily computed to be $F = 1$, giving by substitution

$$u_1(w) = \ln(3) - \frac{3}{2}\ln(2) + \ln(w)$$

for ω_3 and ω_4.

The dynamic programming equation for $t = 1$ is

$$u_0(w) = \max_{F \geqslant 0} E[u_1(w\{1 + F\Delta R_1\}]$$

$$= \max_{F \geqslant 0} \left\{\frac{1}{2}\ln[w(1 + 3F/5)] + \frac{1}{2}\left(\ln(3) - \frac{3}{2}\ln(2) + \ln[w(1 - F/5)\right)\right\}$$

The maximizing value of F is found, as above, to be $F = 5/3$, and substitution gives the value function

$$u_0(v) = J(v) = \ln(v) + \frac{1}{4}\ln(2)$$

In summary, the optimal strategy is to invest $5v/3$ dollars in the risky security at time 0 (since $S_0 = 5$, this means one should buy $v/3$ units or shares), borrowing $2v/3$ dollars from the bank. If $S_1 = 8$, then at time 1 the portfolio will be worth $2v$ dollars; all of this should be held in the bank account until time 2, ending up with $2v$ dollars in both states ω_1 and ω_2. If $S_1 = 4$, then at time 1 the portfolio will be worth

$2v/3$ dollars. In this case it is optimal to invest exactly this sum in the risky security (that is, go long $v/6$ units), taking no position with the bank. Hence one will end up with v and $v/2$ dollars in states ω_3 and ω_4, respectively.

The risk neutral computational approach for the constrained, multiperiod problem is very similar to the approach for the single period problem, with one important exception: the scalar parameter v will now be a predictable stochastic process. As before, the *support function* $\delta(x) : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ of $\mathbb{K}$ is convex and defined by

$$\delta(x) \equiv \sup_{F \in \mathbb{K}} (-Fx')$$

The *effective domain* of δ is the convex cone $\tilde{\mathbb{K}} \equiv \{x \in \mathbb{R}^N : \delta(x) < \infty\}$. It will be assumed that $\mathbb{K}$ is such that δ is continuous on $\tilde{\mathbb{K}}$. In addition, it will be assumed $0 \in \mathbb{K}$ so $\delta \geqslant 0$. Examples of δ and $\tilde{\mathbb{K}}$ are given in section 2.5.

We now introduce a predictable stochastic process $\kappa = \{\kappa(t); t = 1, \dots, T\}$ which is required to satisfy $\kappa(t) \in \tilde{\mathbb{K}}$ for all $t \geqslant 1$. Thus $\kappa(t, \omega)$ will be an N-vector; its nth component will correspond to the nth risky security, as will be shown in a moment. Let $\mathcal{N}$ denote the set of all such processes κ.

We next define an auxiliary market $\mathcal{M}_\kappa$ for each $\kappa \in \mathcal{N}$ by modifying the return processes for the bank account and the risky securities according to

$$r_t \to r_t + \delta(\kappa(t)), \qquad t \geqslant 1$$
$$\Delta R_n(t) \to \Delta R_n(t) + \delta(\kappa(t)) + \kappa_n(t), \qquad n = 1, \dots, N; t \geqslant 1$$

For each such market, let Q_κ denote a corresponding risk neutral probability measure, if one exists. Notice that the market $\mathcal{M}_0$ with $\kappa = 0$ is the same as the original market, since $\delta(0) = 0$. It will usually be the case that Q_κ exists and is unique, in which case a unique Q_κ will also exist for all $\kappa \in \mathcal{N}$ in some neighborhood of $\kappa = 0$, by the assumed continuity of $\delta(\cdot)$.

For the market $\mathcal{M}_\kappa$ and any trading strategy F, whether it is admissible or not, the time T value of the portfolio is

$$(5.35) \qquad V_T^\kappa = v\prod_{t=1}^{T}\left[1 + r_t + \delta(\kappa_t) + \sum_{n=1}^{N} F_n(t)\{\Delta R_n(t) + \kappa_n(t) - r_t\}\right]$$

$$= v\prod_{t=1}^{T}\left[1 + r_t + \sum_{n=1}^{N} F_n(t)\{\Delta R_n(t) - r_t\} + \delta(\kappa_t) + F(t)\kappa_t'\right]$$

where the scalar $F(t)\kappa_t'$ denotes the inner product of the row vector $F(t)$ with the column vector κ_t'.

For each $\kappa \in \mathcal{N}$ we shall be interested in the unconstrained optimal portfolio problem:

$$\text{maximize} \quad Eu(V_T^\kappa)$$
$$\text{subject to} \quad V_0 = v$$

In other words, this is the ordinary optimal portfolio problem for the market $\mathcal{M}_\kappa$. Let $J_\kappa(v)$ denote the corresponding optimal objective value.

Consider an arbitrary market $\mathscr{M}_\kappa$ with $\kappa \neq 0$. As with the single period case, if $F(t) \in \mathbb{K}$ for all t, then by the definition of $\delta(\cdot)$ one has $\delta(\kappa_t) + F(t)\kappa_t' \geqslant 0$ for all ω and t, in which case by (5.35) $V_T^\kappa \geqslant V_T^0$ for all $\omega \in \Omega$. On the other hand, if $F(t) \notin \mathbb{K}$ for some t, then $\delta(\kappa_t) + F(t)\kappa_t' \geqslant 0$ might not hold for all ω and t, in which case one might have $V_T^\kappa < V_T^0$ for some $\omega \in \Omega$. In particular, if $F(t) \notin \mathbb{K}$ for some t, then it is possible to have $Eu(V_T^\kappa) < Eu(V_T^0)$. For this reason, it is possible to have $J_\kappa(v) < J_0(v)$ for some $\kappa \in \mathcal{N}$, that is, the optimal objective value for the unconstrained problem in some of the auxiliary markets may be strictly less than the optimal objective value for the unconstrained problem in the original market.

Now for an arbitrary market $\mathscr{M}_\kappa$ with $\kappa \neq 0$, it is clear that $J_\kappa(v)$, the optimal objective value for the unconstrained problem, is greater than or equal to the optimal objective value for the constrained problem (because if you add constraints, then the optimal objective value, which is being maximized, will not increase).

Meanwhile, suppose $F(t)$ denotes the optimal trading strategy for the original constrained problem, which has optimal objective value $J(v)$. By (5.35) we have $V_T^0 \leqslant V_T^\kappa$, in which case $J(v) = Eu(V_T^0) \leqslant Eu(V_T^\kappa)$. But the right hand side of this inequality will be less than or equal to the optimal objective value for the constrained problem in the market $\mathscr{M}_\kappa$.

Putting together the inequalities of the two preceding paragraphs, we thus have

(5.36) $$J(v) \leqslant J_\kappa(v), \qquad \text{all} \quad \kappa \in \mathcal{N}$$

If this is an equality for some $\kappa \in \mathcal{N}$, then the optimal F corresponding to the right hand side is a candidate as the solution of the original constrained problem, because of the following generalization of principle (2.49):

(5.37) Suppose for some $\hat{\kappa} \in \mathcal{N}$ that F, the optimal trading strategy for the unconstrained portfolio problem in the market $\mathscr{M}_{\hat{\kappa}}$, satisfies

(a) $F \in \mathscr{A}$ (i.e.,$F(t) \in \mathbb{K}$ for all $t \geqslant 1$)

(b) $\delta(\hat{\kappa}(t)) + F(t)\hat{\kappa}'(t) = 0, \qquad \text{all } t \geqslant 1$

Then F is optimal for the constrained problem in the original market $\mathscr{M}_0$ and $J(v) = J_{\hat{\kappa}}(v) \leqslant J_\kappa(v)$ for all $\kappa \in \mathcal{N}$.

To see this, note by expression (5.35) for V_T^κ that W, the attainable wealth under F in the market $\mathscr{M}_{\hat{\kappa}}$, is also the attainable wealth under F in the original market $\mathscr{M}_0$. Since F is feasible for the constrained problem, it follows that $Eu(w) \leqslant J(v)$. But $Eu(w) = J_{\hat{\kappa}}(v)$, so by (5.36) we must have $Eu(w) = J(v) = J_{\hat{\kappa}}(v) \leqslant J_\kappa(v)$ for all $\kappa \in \mathcal{N}$.

In summary, with the risk neutral computational approach you first solve the *dual problem*

$$\begin{array}{c}\text{minimize } J_\kappa(v) \\ \kappa \in \mathcal{N}\end{array}$$

If $\hat{\kappa}$ denotes the optimal solution, then the optimal trading strategy for the unconstrained problem in the market $\mathscr{M}_{\hat{\kappa}}$ will be the candidate for the solution of the constrained problem in the original market $\mathscr{M}_0$. All that remains is to verify that this strategy satisfies the two conditions in (5.37). Perhaps surprisingly, this procedure is usually efficient and successful; an explanation of why this is so is rather difficult and will not be given here.

Example 5.11 (continued) For this single risky security, two-period model with the no short selling constraint $\mathbb{K} = [0, \infty)$, one has

$$\delta(X) = \begin{cases} 0, & x \geqslant 0 \\ \infty, & x < 0 \end{cases}$$

and $\tilde{\mathbb{K}} = [0, \infty)$. The set $\mathcal{N}$ of admissible predictable processes will be all κ of the form

$$\kappa_t(\omega) = \begin{cases} \kappa^5, & t = 1, \\ \kappa^8, & t = 2, \omega = \omega_1, \omega_2 \\ \kappa^4, & t = 2, \omega = \omega_3, \omega_4 \end{cases}$$

where κ^5, κ^8 and κ^4 are non-negative scalars. Since $\delta(x) = 0$ for all $x \in \tilde{\mathbb{K}}$, the interest rate process in the market $\mathscr{M}_\kappa$ is the same as the interest rate process in the market $\mathscr{M}_0$, which is taken to be $r_1 = r_2 = 0$. Hence the bank account process $B_t^\kappa = 1$ for all t and all $\kappa \in \mathcal{N}$.

The return process R_t^κ for the risky security will vary with $\kappa \in \mathcal{N}$. This is shown below, along with the risk neutral probability measure Q_κ and the state price density L_κ:

ω	$\Delta R_1^\kappa(\omega)$	$\Delta R_2^\kappa(\omega)$	$Q_\kappa(\omega)$	$L_\kappa(\omega)$
ω_1	$3/5 + \kappa^5$	$1/8 + \kappa^8$	$\frac{(1-5\kappa^5)(2-8\kappa^8)}{12}$	$\frac{(1-5\kappa^5)(2-8\kappa^8)}{3}$
ω_2	$3/5 + \kappa^5$	$-1/4 + \kappa^8$	$\frac{(1-5\kappa^5)(1+8\kappa^8)}{3}$	$\frac{(1-5\kappa^5)(1+8\kappa^8)}{3}$
ω_3	$-1/5 + \kappa^5$	$1/2 + \kappa^4$	$\frac{(3+5\kappa^5)(1-4\kappa^4)}{12}$	$\frac{(3+5\kappa^5)(1-4\kappa^4)}{3}$
ω_4	$-1/5 + \kappa^5$	$-1/4 + \kappa^4$	$\frac{(3+5\kappa^5)(2+4\kappa^4)}{12}$	$\frac{(3+5\kappa^5)(2+4\kappa^4)}{3}$

The optimal attainable wealth and the value function for the unconstrained problem in the market $\mathscr{M}_\kappa$ are given respectively by

$$W_\kappa = v/L_\kappa = \begin{cases} \frac{3v}{(1-5\kappa^5)(2-8\kappa^8)}, & \omega = \omega_1, \\ \frac{3v}{(1-5\kappa^5)(1+8\kappa^8)}, & \omega = \omega_2, \\ \frac{3v}{(3+5\kappa^5)(1-4\kappa^4)}, & \omega = \omega_3, \\ \frac{3v}{(3+5\kappa^5)(2+4\kappa^8)}, & \omega = \omega_4 \end{cases}$$

and

$$J_\kappa(v) = \ln(v) - E\ln(L_\kappa)$$

$$= \ln(v) - \frac{1}{4}\ln\left(\frac{(1-5\kappa^5)(2-8\kappa^8)}{3}\right) - \frac{1}{4}\ln\left(\frac{(1-5\kappa^5)(1+8\kappa^8)}{3}\right)$$

$$-\frac{1}{4}\ln\left(\frac{(3+5\kappa^5)(1-4\kappa^4)}{3}\right) - \frac{1}{4}\ln\left(\frac{(3+5\kappa^5)(2+4\kappa^4)}{3}\right)$$

Hence the dual problem is:

$$\text{maximize } 2\ln(1-5\kappa^5) + 2\ln(3+5\kappa^5) + \ln(2-8\kappa^8) + \ln(1+8\kappa^8) + \ln(1-4\kappa^4) + \ln(2+4\kappa^4)$$

$$\text{subject to} \quad \kappa^5 \geqslant 0, \quad \kappa^8 \geqslant 0, \quad \kappa^4 \geqslant 0$$

The optimal solution of this is easily found to be

$$\kappa^5 = 0, \quad \kappa^8 = \frac{1}{16}, \quad \kappa^4 = 0$$

Substituting these values in the above expressions for W_κ and $J_\kappa(v)$ gives

$$W_\kappa = \begin{cases} 2v, & \omega = \omega_1, \omega_2 \\ v, & \omega = \omega_3 \\ v/2, & \omega = \omega_4 \end{cases}$$

and $J_\kappa(v) = \ln(v) + \ln(2)/4$ (Note these are the same values that were obtained with dynamic programming). The replicating trading strategy is $F_1 = 5/3; F_2 = 0$ if $\omega = \omega_1, \omega_2$; and $F_2 = 1$ if $\omega = \omega_3, \omega_4$. The two conditions in (5.36) are clearly satisfied, so this must be the optimal strategy for the constrained problem.

Exercise 5.19 For the security model in example 5.11 with $r = 0$ and log utility, suppose short selling is allowed but you cannot borrow money from the bank. Solve this optimal portfolio problem with:

(a) dynamic programming (Hint: show $u_0(v) = \ln(v) + \ln(6/5)$)
(b) the risk neutral computational approach (Hint: show the optimal solution of the dual problem is $\kappa(1) = -\frac{1}{15}$ and $\kappa(2) = 0$).

5.7 Optimal Consumption–Investment With Constraints

It is straightforward to take the optimal consumption/investment model of sections 5.3 and 5.4, add constraints on the admissible trading strategies, and solve the resulting problem with either dynamic programming or a risk neutral computational approach. Indeed, there is no problem extending these ideas to the model studied in section 5.5, where multiperiod consumption/investment is combined with terminal wealth. All this will be explained in this section.

As in section 5.6, the trading strategies are predictable and of the form $F = (F_1, \ldots, F_N)$, with $F_n(t)$ representing the fraction of the time $t-1$ wealth that is invested at that time in security n and held until time t. A non-empty, closed, convex subset $\mathbb{K}$ of $\mathbb{R}^N$ is specified, and it is required that $F(t) \in \mathbb{K}$ for $t = 1, \ldots, T$. A *consumption process* $C = \{C_t; t = 0, \ldots, T\}$ is an adapted, non-negative stochastic process with C_t representing the amount of funds consumed by the investor at time t. A *consumption–investment plan* consists of a pair (C, F), where C is a consumption process and F is a trading strategy.

As usual, the consumption–investment plan (C, F) will be called *self-financing* if no money is added to or withdrawn from the portfolio between times 0 and T, other than the amounts that are consumed. With V_t denoting the value of the portfolio just before any time t transactions or consumption, for self-financing plans it means that

$$V_t = (V_{t-1} - C_{t-1})\left[1 + r_t + \sum_{n=1}^{N} F_n(t)\{\Delta R_n(t) - r_t\}\right] \tag{5.38}$$

for $t = 1, \ldots, T$, where $r_t = (B_t - B_{t-1})/B_{t-1}$ is the interest rate associated with the interval $(t-1, t)$ and $\Delta R_n(t) = R_n(t) - R_n(t-1)$ is the change in the return process for security n. A consumption–investment plan will be called *admissible* if it is self-financing and $V_T \geqslant C_T$. Since consumption processes are non-negative, this implies $V_T \geqslant 0$. Let $\mathscr{A}_v$ denote the set of all admissible consumption–investment plans with initial wealth v.

With the specification of the initial wealth v, the discount parameter satisfying $0 < \alpha \leqslant 1$, and the concave, increasing utility function u, the investor's consumption–investment problem is:

$$\text{maximize } E\left[\sum_{t=0}^{T} \alpha^t u(C_t)\right] \tag{5.39}$$

$$\text{subject to: } (C, F) \in \mathscr{A}_v$$

The corresponding *value function* is

$$J(v) = \sup_{(C,\, F) \in \mathscr{A}_v} E\left[\sum_{t=0}^{T} \alpha^t u(C_t)\right]$$

Note that (5.39) is the same as problem formulation (5.14) in section 5.3, except that implicit is the added requirement that $F(t) \in \mathbb{K}$ for all t.

This problem can be solved by using recursively the dynamic programming functional equation

$$u_{t-1}(w) = \max_{\substack{F \in \mathbb{K} \\ 0 \leqslant c \leqslant w}} \left\{ u(c) + \alpha E\left[u_t\left((w-c)\left\{ 1 + r_t + \sum_{n=1}^{N} F_n(t)\{\Delta R_n(t) - r_t\} \right\} \right) \Big| \mathscr{F}_{t-1} \right] \right\}$$

for $t = 1, 2, \ldots, T$ together with the boundary condition $u_T(w) = u(w)$. Thus $u_t(w)$ represents the maximum expected utility of consumption from time t onwards when you start with initial wealth w at time t and the information available is $\mathscr{F}_t$. Moreover, $J(v) = u_0(v)$. With the common assumption that $u'(0) = \infty$, the explicit constraint $0 \leqslant C$ can be eliminated, thereby simplifying the computations. The dynamic programming approach is illustrated in the following example.

Example 5.12 Consider the two-period, single security model in example 5.1 with interest rate $r_t = 0$, the same situation as in example 5.11. Suppose the utility function $u(c) = -c^{-1}$ and short selling of the risky security is prohibited, so $\mathbb{K} = [0, \infty)$. With $P(\omega_1) = \ldots = P(\omega_4) = 1/4$, the dynamic programming equation for $t = 2$ and $\omega = \omega_1, \omega_2$ is

$$u_1(w) = \max_{\substack{F \geqslant 0 \\ 0 \leqslant c \leqslant w}} \left\{ -\frac{1}{c} - \alpha E\left[\frac{1}{(w-c)(1+F\Delta R_2)} \mid S_1 = 8 \right] \right\}$$

$$= \max_{\substack{F \geqslant 0 \\ 0 \leqslant c \leqslant w}} \left\{ -\frac{1}{c} - \frac{\alpha}{2(w-c)(1+F/8)} - \frac{\alpha}{2(w-c)(1-F/4)} \right\}$$

Computing the partial derivative of the argument with respect to F and setting this equal to zero gives

$$F = \frac{8 - 8\sqrt{2}}{2 + \sqrt{2}} < 0$$

But this violates the short-selling constraint, so take $F = 0$ and substitute this into the above dynamic programming equation, giving

$$u_1(w) = \max_{0 \leqslant c \leqslant w} \left\{ -\frac{1}{c} - \frac{\alpha}{w-c} \right\}$$

Setting the derivative of the argument with respect to c equal to zero, one easily obtains $C = w/(1 + \sqrt{\alpha})$ for the optimal consumption quantity and

$$u_1(w) = -\frac{(1+\sqrt{\alpha})^2}{w}, \qquad \omega = \omega_1, \omega_2$$

for the maximum expected utility.

For $t = 2$ and $\omega = \omega_3, \omega_4$ the dynamic programming equation is

$$u_1(w) = \max_{\substack{F \geqslant 0 \\ 0 \leqslant c \leqslant w}} \left\{ -\frac{1}{c} - \frac{\alpha}{2(w-c)(1+F/2)} - \frac{\alpha}{2(w-c)(1-F/4)} \right\}$$

Setting to zero the partial derivative with respect to F leads to

$$F = \frac{4\sqrt{2} - 4}{2 + \sqrt{2}} > 0$$

which satisfies the short-selling constraint. Substituting this into the dynamic programming equation yields

$$u_1(w) = \max_{0 \leqslant c \leqslant w} \left\{ -\frac{1}{c} - \frac{\alpha(1+\sqrt{2})^2}{6(w-c)} \right\}$$

Hence the optimal consumption quantity is $C = w/(1 + \sqrt{\alpha_1})$ and the maximum expected utility is

$$u_1(w) = -\frac{(1+\sqrt{\alpha_1})^2}{w} \qquad \omega = \omega_3, \omega_4$$

where for convenience the new parameter

$$\alpha_1 = \frac{\alpha(1+\sqrt{2})^2}{6}$$

has been introduced.

For $t = 1$ the dynamic programming equation is

$$u_0(w) = \max_{\substack{F \geqslant 0 \\ 0 \leqslant c \leqslant w}} \left\{ -\frac{1}{c} - \frac{\alpha(1+\sqrt{\alpha})^2}{2(w-c)(1+3F/5)} - \frac{\alpha(1+\sqrt{\alpha_1})^2}{2(w-c)(1-F/5)} \right\}$$

Setting the partial derivative of the argument with respect to F equal to zero leads to

$$F = \frac{5[\sqrt{3}(1+\sqrt{\alpha}) - 1 - \sqrt{\alpha_1}]}{\sqrt{3}(1+\sqrt{\alpha}) + 3 + 3\sqrt{\alpha_1}}$$

This is easily verified to be positive, so the short-selling constraint is satisfied. Substituting this back into the dynamic programming equation gives

$$u_0(w) = \max_{0 \leqslant c \leqslant w} \left\{ -\frac{1}{c} - \frac{\alpha_0}{w-c} \right\}$$

where for convenience the new parameter

$$\alpha_0 = \frac{\alpha}{8}\left[1 + \sqrt{\alpha} + \sqrt{3} + \sqrt{3\alpha_1}\right]^2$$

has been introduced. Hence $C = w/(1 + \sqrt{\alpha_0})$ is the optimal time-0 consumption quantity and

$$J(v) = u_0(v) = -\frac{(1 + \sqrt{\alpha_0})^2}{v}$$

is the value function for the original problem.

The risk neutral computational approach for the multiperiod consumption–investment problem is essentially the same as for the multiperiod optimal portfolio problem. The *support function* $\delta(x) : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ of $-\mathbb{K}$ is defined by

$$\delta(x) \equiv \sup_{F \in \mathbb{K}} (-Fx)$$

the *effective domain* of which is the convex cone $\tilde{\mathbb{K}} \equiv \{x \in \mathbb{R}^N : \delta(x) < \infty\}$. It will be assumed that $\mathbb{K}$ is such that δ is continuous on $\tilde{\mathbb{K}}$ with $0 \in \mathbb{K}$.

Let $\mathcal{N}$ denote the set of all predictable stochastic processes $\kappa = \{\kappa(t) : t = 1, \ldots, T\}$ satisfying $\kappa(t) \in \tilde{\mathbb{K}}$ for all t. For each $\kappa \in \mathcal{N}$ one defines an auxiliary market $\mathcal{M}_\kappa$ by modifying the return processes for the bank account and the risky securities according to

$$\begin{aligned} r_t &\to r_t + \delta(\kappa(t)) & t &\geqslant 1 \\ \Delta R_n(t) &\to \Delta R_n(t) + \delta(\kappa(t)) + \kappa_n(t) & n &= 1, \ldots, N; t \geqslant 1 \end{aligned}$$

For each such market, let Q_κ denote a corresponding risk neutral probability measure. It will usually be the case that $Q = Q_0$ exists and is unique, in which case a unique Q_κ will also exist for all $\kappa \in \mathcal{N}$ in some neighbourhood of $\kappa = 0$.

For the market $\mathcal{M}_\kappa$ and any self-financing consumption plan (C, F), whether it is admissible or not, the self-financing equation (5.38) becomes

(5.40)

$$V_t = (V_{t-1} - C_{t-1})\left[1 + r_t + \sum_{n=1}^{N} F_n(t)\{\Delta R_n(t) - r_T\} + \delta(\kappa_t) + F(t)\kappa_t'\right]$$

where the scalar $F(t)\kappa'(t)$ denotes the inner product of the row vector $F(t)$ with the column vector $\kappa'(t) = \kappa_t'$. Notice that the market $\mathcal{M}_0$ with $\kappa = 0$ is the same as the original market.

For each $\kappa \in \mathcal{N}$ we shall be interested in the unconstrained consumption–investment problem:

(5.41) maximize $\quad E\left[\sum_{t=0}^{T} \alpha^t u(C_t)\right]$

subject to: $\quad V_0 = v, \quad V_T \geqslant C_T$

$V,\ C,\ F$ satisfy (5.40) for $t \geqslant 1$

In other words, this is the ordinary consumption–investment problem for the market $\mathcal{M}_\kappa$, with no special constraints on the values of F. Let $J_\kappa(v)$ denote the corresponding optimal objective value.

As we saw for constrained optimal portfolio problems in section 5.6,

(5.42) $\quad J(v) \leqslant J_\kappa(v), \quad$ all $\kappa \in \mathcal{N}$

that is, the optimal objective value for the original constrained problem is less than or equal to the optimal objective value for the unconstrained problem in market $\mathcal{M}_\kappa$, provided $\kappa \in \mathcal{N}$. This is because if (C, F) is feasible for (5.39), then (5.40) together with $\delta(\kappa_t) + F(t)\kappa_t' \geqslant 0$ imply (C, F) is also feasible for (5.41). Hence if (5.42) is an equality for some $\kappa \in \mathcal{N}$, then the optimal (C, F) corresponding to this v is a candidate for the solution of the original constrained problem, because of the following counterpart of (5.37) (which is true by exactly the same kind of logic):

(5.43) Suppose for some $\hat{\kappa} \in \mathcal{N}$ that (C, F), the optimal consumption–investment plan for the unconstrained consumption–investment problem in the market $\mathcal{M}_{\hat{\kappa}}$, satisfies

(a) $(C, F) \in \mathcal{A}_\kappa$

(b) $\delta(\hat{\kappa}_t) + F(t)\hat{\kappa}_t' \geqslant 0, \quad$ all $t \geqslant 1$

Then (C, F) is optimal for the original constrained consumption–investment problem, and $J(v) = J_{\hat{\kappa}}(v) \leqslant J_\kappa(v)$ for all $\kappa \in \mathcal{N}$.

In summary, to solve the constrained consumption–investment problem with the risk neutral computational approach, you first solve the dual problem of minimizing the right hand side of (5.42) over $\kappa \in \mathcal{N}$. This will be illustrated in the following example.

Example 5.12 (continued) Since $\mathbb{K} = [0, \infty)$ and the original markets are the same, the auxiliary markets $\mathcal{M}_\kappa$ will be the same as in example 5.11. In particular, the set $\mathcal{N}$, the notation for the processes κ, the return processes R^κ, and the risk neutral probability measures Q_κ are given in section 5.6.

In order to solve the unconstrained problem (5.41) for each $\kappa \in \mathcal{N}$, it is necessary to introduce for each κ the process N^κ corresponding to N of (5.22); this is given below (recall $B_t^\kappa = 1$ for all t):

ω	$N_0^\nu(\omega)$	$N_1^\nu(\omega)$	$N_2^\nu(\omega)$
ω_1	1	$(1-5\kappa^5)/2$	$(1-5\kappa^5)(2-8\kappa^8)/3$
ω_2	1	$(1-5\kappa^5)/2$	$(1-5\kappa^5)(1+8\kappa^8)/3$
ω_3	1	$(3+5\kappa^5)/2$	$(3+5\kappa^5)(1-4\kappa^4)/3$
ω_4	1	$(3+5\kappa^5)/2$	$(3+5\kappa^5)(2+4\kappa^4)/3$

By exercise 5.14, the optimal objective value for the unconstrained problem (5.41) is given by

$$J_\kappa(v) = -\frac{\Delta^2}{v}$$

where

$$\Delta = \sum_{t=0}^{2} \alpha^{t/2} E[\sqrt{N_t^\kappa}]$$

Hence the dual problem amounts to maximizing Δ^2 over all $\kappa \in \mathcal{N}$, that is, over all non-negative, predictable κ. With a little bit of work, the solution is found to be

$$\hat\kappa^5 = 0, \quad \hat\kappa^8 = \frac{1}{16}, \quad \hat\kappa^8 = 0$$

giving the corresponding value

$$\Delta = 1 + \frac{\sqrt{\alpha}}{2\sqrt{2}}\left[1+\sqrt{3}\right] + \frac{\alpha}{4}\left[1+2\sqrt{2}\right]$$

Note that this value of Δ gives, after a little algebra, $J_{\hat\kappa}(v) = -(1+\sqrt{\alpha_0})^2/v$, the same as the optimal objective value for the original constrained consumption–investment problem, as computed earlier with dynamic programming.

Since $N_1^\kappa(\omega_1) = N_1^\kappa(\omega_2) = 1/2$, $N_1^\kappa(\omega_3) = N_1^\kappa(\omega_4) = 3/2$, $N_2^\kappa(\omega_1) = N_2^\kappa(\omega_2) = 1/2$, $N_2^\kappa(\omega_3) = 1$ and $N_2^\kappa(\omega_4) = 2$, exercise 5.14 also gives $C_1(\omega_1) = C_1(\omega_2) = v\sqrt{2\alpha}/\Delta$, $C_1(\omega_3) = C_1(\omega_4) = v\sqrt{2\alpha/3}/\Delta$, $C_2(\omega_1) = C_2(\omega_2) = v\alpha\sqrt{2}/\Delta$, $C_2(\omega_3) = v\alpha/\Delta$, and $C_2(\omega_4) = v\alpha/(\Delta\sqrt{2})$.

To compute the replicating trading strategy, we begin by noting that $C_2(\omega_1) = C_2(\omega_2)$ implies $F_2(\omega_1) = F_2(\omega_2) = 0$ for the optimal fraction of the available funds that are invested in the risky security at time 1. It follows that

$$V_1(\omega) = C_1(\omega) + C_2(\omega) = \frac{v}{\Delta}\left[\alpha\sqrt{2} + \sqrt{2\alpha}\right], \qquad \omega = \omega_1, \omega_2$$

Next, solving

$$(V_1 - C_1)(1 + F/2) = v\alpha/\Delta$$
$$(V_1 - C_1)(1 - F/4) = v\alpha/(\Delta\sqrt{2})$$

gives

$$F_2(\omega_3) = F_2(\omega_4) = \frac{4(\sqrt{2}-1)}{2+\sqrt{2}}$$

and

$$V_1(\omega_3) = V_1(\omega_4) = \frac{v}{\Delta}\left[\frac{\alpha(2+\sqrt{2})}{3\sqrt{2}} + \frac{\sqrt{2\alpha}}{\sqrt{3}}\right]$$

Finally, solving

$$(V_0 - C_0)(1 + 3F/5) = \frac{v}{\Delta}\left[\alpha\sqrt{2} + \sqrt{2\alpha}\right]$$
$$(V_0 - C_0)(1 - F/5) = \frac{v}{\Delta}\left[\frac{\alpha(2+\sqrt{2})}{3\sqrt{2}} + \frac{\sqrt{2\alpha}}{\sqrt{3}}\right]$$

gives the same value for F_1 as was obtained with dynamic programming.

Needless to say, the preceding ideas can be combined with the ideas of section 5.6 to solve the section 5.5 kind of problem where there is utility from both consumption and terminal wealth. In particular, you can use either dynamic programming (see exercise 5.20) or the risk neutral computational approach that will now be explained.

With $\mathcal{A}_v$ the set of admissible consumption–investment plans with initial wealth v and F satisfying the constraint $\mathbb{K}$, we are interested in the value function

$$J(v) = \max_{(C,\,F)\in\mathcal{A}_v} E\left[\sum_{t=0}^{T} \alpha^t u_c(C_t) + \alpha^T u_p(V_T - C_T)\right]$$

where u_c and u_p are two specified utility functions, as in section 5.5. This is the same as (5.28), and this corresponds to exactly the same optimization problem as in section 5.5, only now the constraint $F(t) \in \mathbb{K}$ is implicit.

We now define the auxiliary markets $\mathcal{M}_\kappa$ exactly the same as for the constrained optimal portfolio problem of section 5.6 and as the constrained consumption–investment problem studied earlier in this section. The set $\mathcal{N}$ of predictable processes κ, the modified return processes, and so forth will depend on $\mathbb{K}$ and the original security model, but they will be independent of the utility functions and whether u_c or u_p equals zero.

Next, for each $\kappa \in \mathcal{N}$ we wish to consider the "unconstrained" optimization problem:

$$\text{maximize} \quad E\left[\sum_{t=0}^{T} \alpha^t u(C_t) + \alpha^T u_p(V_T - C_T)\right]$$
$$\text{subject to:} \quad V_0 = v, \quad V_T \geqslant C_T$$
$$V,\ C,\ F \ \text{satisfy (5.40) for t} \geqslant 1$$

This is the same as (5.41), only now there is the additional term associated with the utility of terminal wealth. In particular, this is the same as the optimization problem of section 5.5 for the market $\mathcal{M}_\kappa$, with no special constraints on the values of F. Thus $J_\kappa(v)$, the corresponding optimal objective value, can be computed with the methods of that section.

Finally, we solve the dual problem of minimizing $J_\kappa(v)$ over $\kappa \in \mathcal{N}$. If $\hat{\kappa} \in \mathcal{N}$ is the minimizing value and $(\hat{C}, \hat{F})$ is the corresponding optimal consumption–investment plan, then it only remains to check the two conditions in (5.43) in order to conclude that $(\hat{C}, \hat{F})$ is also the optimal solution for the original constrained optimization problem.

Example 5.13 Consider the two-period, single security model in example 5.1 with interest rate $r_t = 0$. This is the same situation as in examples 5.11 and 5.12, except now it is assumed that $u_c(c) = \ln(c)$ and $u_p(w) = \ln(w)$. The auxiliary markets $\mathcal{M}_\kappa$, including the processes N^κ, are the same as in example 5.12.

In view of section 5.5, the optimal solution for the unconstrained problem is of the form $C_t = \alpha^t/(\lambda N_t^\kappa)$ and $V_2 = 2\alpha^2/(\lambda N_2^\kappa)$, where the process N^κ depends on the underlying market $\mathcal{M}_\kappa$. Since $f(\lambda) = (1 + \alpha + 2\alpha^2)/\lambda$, it follows that $\lambda = g(v) = (1 + \alpha + 2\alpha^2)/v$ and

$$C_t = \frac{\alpha^t v}{(1+\alpha+2\alpha^2)N_t^\kappa} \qquad V_2 = \frac{2\alpha^2 v}{(1+\alpha+2\alpha^2)N_2^\kappa}$$

Moreover, since

$$K(\lambda) = -(1+\alpha+2\alpha^2)\ln\lambda + (\alpha+4\alpha^2)\ln\alpha - \alpha E \ln N_1^\kappa - 2\alpha^2 E \ln N_2^\kappa$$

it follows from $J_\kappa(v) = K(g(v))$ that

$$J_\kappa(v) = (1+\alpha+2\alpha^2)\ln\left(\frac{v}{1+\alpha+2\alpha^2}\right) + (\alpha+4\alpha^2)\ln\alpha - \alpha E \ln N_1^\kappa - 2\alpha^2 E \ln N_2^\kappa$$

Now upon substituting N^κ (see example 5.12), it is apparent the dual problem amounts to maximizing

$$(\alpha/2+\alpha^2)\left[\ln(1-5\kappa^5) + \ln(3+5\kappa^5)\right] - \alpha\ln(2) - 2\alpha^2\ln(3) + \frac{\alpha^2}{2}\left[\ln(2-8\kappa^8) + \ln(1+8\kappa^8) + \ln(1-4\kappa^4) + \ln(2+4\kappa^4)\right]$$

over non-negative values of the scalars κ^4, κ^5 and κ^8. The optimal solution is easily found to be $\kappa^4 = \kappa^5 = 0$, $\kappa^8 = 1/16$. The corresponding process N^κ is

$$N_1^\kappa(\omega) = \begin{cases} 1/2, & \omega_1, \omega_2 \\ 3/2, & \omega_3, \omega_4 \end{cases} \qquad N_2^\kappa(\omega) = \begin{cases} 1/2, & \omega_1, \omega_2 \\ 1, & \omega_3 \\ 2, & \omega_4 \end{cases}$$

in which case $C_0 = v/(1 + \alpha + 2\alpha^2)$,

$$C_1(\omega) = \begin{cases} \dfrac{2\alpha v}{1+\alpha+2\alpha^2}, & \omega_1, \omega_2 \\ \dfrac{2\alpha v/3}{1+\alpha+2\alpha^2}, & \omega_3, \omega_4 \end{cases} \qquad C_2(\omega) = \begin{cases} \dfrac{2\alpha^2 v}{1+\alpha+2\alpha^2}, & \omega_1, \omega_2 \\ \dfrac{\alpha^2 v}{1+\alpha+2\alpha^2}, & \omega_3 \\ \dfrac{\alpha^2 v/2}{1+\alpha+2\alpha^2}, & \omega_4 \end{cases}$$

and $V_2 = 2C_2$. After computing the replicating trading strategy F, one verifies that the two conditions in (5.43) are satisfied, which means this consumption–investment plan must be optimal for the constrained problem. The optimal objective value $J(v)$ is

$$(1+\alpha+2\alpha^2)\ln\left(\frac{v}{1+\alpha+2\alpha^2}\right) + (\alpha+4\alpha^2)\ln\alpha + (\alpha+\alpha^2/2)\ln 2 - \frac{\alpha}{2}\ln 3$$

Exercise 5.20 Use dynamic programming to compute the optimal trading strategy F for the problem in example 5.13. Show that $F_1 = 5/3$, $F_2(\omega) = 0$ when $\omega = \omega_1, \omega_2$, and $F_2(\omega) = 1$ when $\omega = \omega_3, \omega_4$. Verify that you get the same value for $J(v)$ as in example 5.13.

5.8 Portfolio Optimization in Incomplete Markets

In view of section 2.6 and the themes of this chapter, there are three principal approaches to solving optimal portfolio and consumption–investment problems when the underlying securities market is incomplete: dynamic programming, the convex optimization approach using Lagrange multipliers as in sections 5.2 and 5.4, only with multiple constraints and multiple risk neutral probability measures, and an approach featuring augmented fictitious securities coupled with constraints on the trading strategies. These three approaches are straightforward extensions of principles studied earlier in this chapter, so the emphasis will be on the presentation of an example, with redundant explanations kept to a minimum.

The dynamic programming approach for incomplete markets is exactly the same as for complete markets. The computational requirements depend primarily on the number of periods, the number of nodes in the information tree, and the number of securities, but not on whether the market is complete. It is largely a matter of taste whether one chooses the trading strategy in terms of the units held of each security (i.e., H) or in terms of the fractions of wealth allocated among the securities (i.e., F).

Example 5.14 The incomplete securities market model is the same as example 4.10, namely, $K = 5$, $N = 1$, $r = 0$, the filtration $\mathcal{F}_t$ is generated by the price process S, and the specifications of S, the underlying

probability measure P, and the family of risk neutral probability measures Q are:

ω	S_0	S_1	S_2	$\Delta R(1)$	$\Delta R(2)$	P	Q
ω_1	5	8	9	3/5	1/8	1/5	$q/4$
ω_2	5	8	7	3/5	−1/8	1/5	$(2-3q)/4$
ω_3	5	8	6	3/5	−1/4	1/5	$(2q-1)/4$
ω_4	5	4	6	−1/5	1/2	1/5	1/4
ω_5	5	4	3	−1/5	−1/4	1/5	1/2

The parameter q here is any scalar satisfying $1/2 < q < 2/3$. The objective will be to maximize expected utility of terminal wealth given the log utility function $u(w) = \ln(w)$.

With trading strategies expressed as the fraction F of wealth held in the risky security, the dynamic programming functional equation is:

$$u_t(w) = \max_F \; E[u_{t+1}(w\{1 + F\Delta R(t+1)\})|\mathscr{F}_t]$$

Taking $u_2(w) = \ln(w)$ when $t = 1$ and $\omega = \omega_1$, ω_2 or ω_3 this is

$$u_t(w) = \max_F \; \frac{1}{3}[\ln(w\{1+F/8\}) + \ln(w\{1-F/8\}) + \ln(w\{1-F/4\})]$$

Notice this utility function forces the time $t = 2$ wealth to be positive, so one has the implicit constraint $-8 < F < 4$. Computing the derivative of the argument with respect to F, the necessary condition leads to the quadratic equation $3F^2 - 8F - 64 = 0$. This equation has two roots, but only one falls in the interval $(-8, 4)$, so

$$F = \frac{4}{3}(1 - \sqrt{13}) \cong -3.4741$$

Substituting this back into the dynamic programming equation eventually leads to $u_1(w) = \ln(\mathrm{w}) + k_8$, where for convenience I have introduced the scalar

$$k_8 = \frac{1}{3}\ln\left[\left(7-\sqrt{13}\right)\left(5+\sqrt{13}\right)\left(4+2\sqrt{13}\right)\right] - \ln 6 \cong 0.0703$$

In a similar fashion one computes for $t = 1$ and $\omega = \omega_4$ or ω_5 the optimal fraction $F = 1$ and $u_1(w) = \ln(w) + k_4$, where k_4 is another new scalar given by $k_4 = \ln(3) - (3/2)\ln(2) \cong 0.0589$.

Turning to the dynamic programming equation for $t = 0$, one has

$$u_0(w) = \max_F \left\{\frac{3}{5}[\ln(w\{1+3F/5\}) + k_8] + \frac{2}{5}[\ln(w\{1-F/5\}) + k_4]\right\}$$

It follows in the usual manner that the optimal fraction is $F = 7/3$ and that the optimal value function $J(v) = u_0(v) = \ln(v) + k_5$ where

$$k_5 = \frac{1}{5}\ln(3) + \frac{12}{5}\ln(2) - \ln(5) + \frac{3}{5}k_8 + \frac{2}{5}k_4 \cong 0.3397$$

Starting with v dollars and implementing the optimal values of F that were just computed, one ends up with the terminal wealth

$$V_2 = \begin{cases} 2v(7-\sqrt{13})/5, & \omega = \omega_1 \\ 2v(5-\sqrt{13})/5, & \omega = \omega_2 \\ 4v(2+\sqrt{13})/5, & \omega = \omega_3 \\ 4v/5, & \omega = \omega_4 \\ 2v/5, & \omega = \omega_5 \end{cases}$$

The second computational approach, which for convenience will be called the Lagrange multiplier approach, is essentially the same as for single period models (see section 2.6). One starts with the standard convex optimization problem (2.50), which now features two or more constraints. Each constraint corresponds to a risk neutral probability measure, and there are enough such risk neutral probability measures to form a "basis" for the set of all risk neutral probability measures. Since there will be one Lagrange multiplier corresponding to each constraint, the amount of computational work will go up, compared to a model for complete markets. This is because a system of equations (one equation for each constraint) will need to be solved to compute the correct value of these Lagrange multipliers.

Example 5.14 (continued) Two risk neutral probability measures suffice to comprise a basis for the set of risk neutral probability measures. Indeed, one can take, as I shall, the extreme points of the closure of this set, corresponding to the parameter values $q = 1/2$ and $q = 2/3$:

$$Q(1) = (1/8,\ 1/8,\ 0,\ 1/4,\ 1/2) \quad Q(2) = (1/6,\ 0,\ 1/12,\ 1/4,\ 1/2)$$

The corresponding state price vectors are:

$$L_1 = (5/8,\ 5/8,\ 0,\ 5/4,\ 5/2) \quad L_2 = (5/6,\ 0,\ 5/12,\ 5/4,\ 5/2)$$

The optimization problem we need to solve is:

$$\begin{aligned} \text{maximize} \quad & E[\ln(W)] \\ \text{subject to} \quad & E_{Q(1)}[W] = v \\ & E_{Q(2)}[W] = v \end{aligned}$$

Proceeding in the usual way, we introduce the Lagrange multipliers λ_1 and λ_2 and use the inverse of the marginal utility function to derive the following expression for the terminal (time $t = 2$) wealth:

$$W(\omega) = \frac{1}{\lambda_1 L_1(\omega) + \lambda_2 L_2(\omega)}$$

The correct values of the Lagrange multipliers are obtained by solving the system:

$$E_{Q(1)}\left[\frac{1}{\lambda_1 L_1 + \lambda_2 L_2}\right] = v$$

$$E_{Q(1)}\left[\frac{1}{\lambda_1 L_1 + \lambda_2 L_2}\right] = v$$

The solution is:

$$\lambda_1 = \frac{4}{v(5+\sqrt{13})} \qquad \lambda_2 = \frac{3}{v(2+\sqrt{13})}$$

Substitution in the above expression for $W(\omega)$ gives the same values as were computed earlier with dynamic programming.

The basic idea of the third approach is the same as with the single period model of section 2.6: first add one or more fictitious securities in such a way as to make the risk neutral probability measure unique. Then solve the portfolio optimization problem with the constraint that precludes trading in the fictitious securities.

To add the fictitious securities, a good approach is to work with return processes and the information tree, constructing the fictitious securities by proceeding from one node to the next. If the conditional risk neutral probabilities at a node are not unique, then add one or more (one-period) fictitious securities to that node so as to make these conditional probabilities unique. The analysis at any one node is equivalent to that for a single period problem.

As one proceeds through the tree, it may turn out that different nodes require the addition of different numbers of fictitious securities. In fact, some nodes may not require any. So after the initial pass through the network, note the maximum number of additional fictitious securities that are required. This becomes the number needed to add to the overall multiperiod problem. It is then necessary to go back to the nodes where less than this maximum number were defined and add some more (one-period) fictitious securities, as necessary, so as to bring the total number up to the maximum. In order to preserve the arbitrage-free property of the securities market, these additional (one-period) securities will need to be 'locally redundant,' that is, linear combinations of the (one-period) securities that are already specified at that same node. For convenience, these additional (one-period) securities that are defined in this 'topping-up' phase should be linear combinations of the original securities but not of any fictitious (one-period) securities that may have been defined during the initial pass through the information tree.

At this stage the same number of (one-period) fictitious securities will have been defined, in terms of their return processes, at each node of the information tree. There will also be a unique risk neutral conditional probability measure at each node. By specifying their time $t = 0$ values, the full specifications of all the (multiperiod) fictitious securities can now be synthesized. And the unique risk neutral probability measure for the overall multiperiod model can readily be computed.

Suppose there is a total of N securities, with the original ones indexed $n = 1, 2, \ldots, \hat{n}$ and the fictitious ones indexed $n = \hat{n}+1, \ldots, N$. I would now like to introduce a constraint of the form

$$\mathbb{K} = \{F \in \mathbb{R}^N : F_{\hat{n}+1} = \ldots = F_N = 0\}$$

In view of section 5.6, this would imply taking

$$\delta(x) = \sup_{F \in \mathbb{K}} (-Fx') = \begin{cases} 0, & x_1 = \ldots = x_{\hat{n}} = 0 \\ \infty, & \text{otherwise} \end{cases}$$

$\tilde{\mathbb{K}} = \{x \in \mathbb{R}^N : x_1 = \ldots = x_{\hat{n}} = 0\}$, and $\mathcal{N}$ the set of all predictable processes κ satisfying $\kappa(t) \in \tilde{\mathbb{K}}$ for all t. Moreover, the return processes for the interest rate and original securities would remain unchanged in the auxiliary market $\mathcal{M}_\kappa$, while the return processes for the fictitious securities would be given by $\Delta R_n(t) + \kappa_n(t)$, where R_n is its return process in the original market.

However, this may lead to a complication. If you look at a node of the information tree where there are "locally redundant" fictitious securities, you may find that there does not exist a conditional risk neutral probability measure for that node in the auxiliary market $\mathcal{M}_\kappa$, because for non-zero κ the securities have ceased to be locally redundant. Indeed, arbitrage opportunities may now exist in every auxiliary market, which means the risk neutral computational procedure for solving the constrained optimization problem breaks down.

The way around this difficulty is to realize that there is no harm in trading a fictitious security if at a particular moment in time it is "locally redundant." In other words, if corresponding to a particular node in the information tree the one-period return for a fictitious security is a linear combination of the returns for the original securities, then we can relax the constraint which precludes trading in that security in that (one-period) circumstance. Hence the sets $\mathbb{K}$ and $\tilde{\mathbb{K}}$ should be defined for each node in the information tree, as these may vary across the network. The set $\mathbb{K}$ at a node should stipulate that $F_n = 0$ if and only if security n is fictitious but it is not "locally redundant" at that node. It follows that for $\mathcal{N}$ one should take all predictable processes κ such that $\kappa(t, \omega)$ is an element of the set $\mathbb{K}$ which corresponds to (t, ω).

As usual, an auxiliary market $\mathcal{M}_\kappa$ will be defined for each $\kappa \in \mathcal{N}$. The return processes for the interest rate and the original securities will remain the same, but in the auxiliary market $\mathcal{M}_\kappa$ the return process for fictitious security n will be given by $\Delta R_n(t) + \kappa_n(t)$, where R_n is its original return process. Now if fictitious security n is locally redundant at a node in the information tree, the corresponding value of $\kappa_n(t)$ will be zero, a unique conditional risk neutral probability measure will exist at that node, and the risk neutral computational approach for solving this constrained optimiza-

tion problem can proceed without difficulty. This all will be illustrated in the following example.

Example 5.14 (continued) We now need to let S_1 denote the original risky security. The conditional risk neutral probability measures for the nodes where $S_1(0) = 5$ and $S_1(1) = 4$ are unique, but this is not true for the node where $S_1(1) = 8$. There we need to add one fictitious security, which we denote security $n = 2$ and take $\Delta R_2(2, \omega_1) = 1/16$, $\Delta R_2(2, \omega_2) = 0$, and $\Delta R_2(2, \omega_3) = -3/16$. This results in the conditional risk neutral probability measure (0.6, 0.2, 0.2), which is a special case of what exists with the original incomplete model. We take $\Delta R_2 = \Delta R_1$ at the other two nodes, so choosing $S_2(0) = 10$, the full specifications of fictitious security S_2 are listed below. The unique risk neutral probability measure Q is also provided.

ω	$S_2(0)$	$S_2(1)$	$S_2(2)$	$\Delta R_2(1)$	$\Delta R_2(2)$	Q
ω_1	10	16	17	3/5	1/16	0.15
ω_2	10	16	16	3/5	0	0.05
ω_3	10	16	13	3/5	−3/16	0.05
ω_4	10	8	12	−1/5	1/2	0.25
ω_5	10	8	6	−1/5	−1/4	0.50

There are no constraints at the nodes where $S_1(0) = 5$ and $S_1(1) = 4$, so there we take $\mathbb{K} = \{x \in \mathbb{R}^2\}$ and $\tilde{\mathbb{K}} = \{(0, 0)\}$. At the node where $S_1(1) = 8$ we want to preclude trading in fictitious security S_2 so there we have $\mathbb{K} = \{x \in \mathbb{R}^2 : F_2 = 0\}$ and $\tilde{\mathbb{K}} = \{x \in \mathbb{R}^2 : x_1 = 0\}$. Hence $\mathcal{N}$ consists of all the predictable processes κ of the form

$$\kappa(t, \omega) = \begin{cases} (0, \tilde{\kappa}), & t = 2, \quad \omega = \omega_1, \omega_2 \\ (0, 0), & \text{otherwise} \end{cases}$$

where $\tilde{\kappa}$ is a scalar.

The next step is to compute the risk neutral probability measure for the market $\mathcal{M}_\kappa$. We have

$$\Delta R_2^\kappa(2, \omega) = \begin{cases} 1/16 + \tilde{\kappa}, & \omega = \omega_1 \\ \tilde{\kappa}, & \omega = \omega_2 \\ -3/16 + \tilde{\kappa}, & \omega = \omega_3 \end{cases}$$

whereas $\Delta R_1^\kappa(2) = \Delta R_2(2)$ for all $\kappa \in \mathcal{N}$. Thus using the two equations $E_{Q_\kappa}[\Delta R_n^\kappa(2) | S_1(1) = 8]$ for $n = 1$ and 2 we solve for and obtain the conditional risk neutral probability measure at the node where $S_1(1) = 8$:

$$\left(\frac{3 + 16\tilde{\kappa}}{5}, \frac{1 - 48\tilde{\kappa}}{5}, \frac{1 + 32\tilde{\kappa}}{5} \right)$$

The conditional risk neutral probability measures at the other two nodes are the same as for the original market $\mathcal{M}_0$, so the risk neutral probability measure for the market $\mathcal{M}_\kappa$ must be:

$$\left(\frac{3 + 16\tilde{\kappa}}{20}, \frac{1 - 48\tilde{\kappa}}{20}, \frac{1 + 32\tilde{\kappa}}{20}, \frac{1}{4}, \frac{1}{2} \right)$$

The next step is to solve the dual problem of minimizing $J_\kappa(v)$ over $\kappa \in \mathcal{N}$. Since with log utility for unconstrained problems $J_\kappa(v) = \ln(v) - E\ln(Q_\kappa) + E\ln(P)$, this amounts to maximizing $E\ln(Q_\kappa)$ over $\tilde{\kappa} \in \mathbb{R}$. Writing this out and differentiating leads to the necessary condition:

$$\frac{16}{3 + 16\tilde{\kappa}} - \frac{48}{1 - 48\tilde{\kappa}} + \frac{32}{1 + 32\tilde{\kappa}} = 0$$

The solution is found to be

$$\tilde{\nu} = \frac{3\sqrt{13} - 11}{16(7 - \sqrt{13})}$$

With log utility for unconstrained problems the optimal terminal wealth is $V_2^\kappa = vP/Q_\kappa$, so substituting $\tilde{\kappa}$ one obtains the final result. This optimal terminal wealth for the constrained problem is equal to the optimal terminal wealth for the original incomplete market model, a terminal wealth that we computed earlier with two other approaches.

It should be clear that multiperiod consumption – investment problems for incomplete markets can also be solved with the same three approaches. The applications of these approaches are so similar to what has already been done that the details will be left to the reader.

Exercise 5.21 Solve the consumption – investment problem for the model in example 5.14, assuming log utility (i.e., $u(c) = \ln(c)$) and a discount parameter α taking a general non-negative value less than or equal to one. Derive the optimal trading strategy, the optimal consumption process, and the optimal value function $J(v)$. In particular, show that the optimal consumption quantity at time $t = 2$ is $2\alpha^2 vX(\omega)/[5(1 + \alpha + \alpha^2)]$, where the random variable X is equal to $7 - \sqrt{13}$, $5 + \sqrt{13}$, $4 + 2\sqrt{13}$, 2, and 1 in states ω_1, ω_2, ω_3, ω_4, and ω_5, respectively.

(a) Use the dynamic programming approach.
(b) Use the Lagrange multiplier approach.
(c) Use the fictitious securities approach.

6 Bonds and Interest Rate Derivatives

6.1 The Basic Term Structure Model

Although the securities market models discussed in the preceding chapters are very general in terms of the kinds of securities that can be modeled, the securities being modeled were usually thought of as equities such as common stocks. However, the case where fixed income securities such as bonds are included among the risky securities is so important that it will be the subject of this entire chapter. The securities market model for this situation is called a *term structure model.*

For a securities market model to be a term structure model, three things are required. First, it must be a multiperiod model. Second, the interest rate r must be a strictly positive, predictable process, so that the interest rate r_t for borrowing and lending over the period $(t-1, t]$ is known at time $t-1$. As usual, $B_0 = 1$ and $r_t = (B_t - B_{t-1})/B_{t-1}$ for $t = 1, \ldots, T$ so the assumption $r_t > 0$ means that the bank account B is strictly increasing in time. Since the term structure model will feature several interest rates, r will be called the *spot* interest rate as well as (although this is misleading, since r can be random) the *riskless* interest rate.

Thirdly, and most important, included among the risky securities is a collection of what are called *zero coupon* or *discount* bonds. Defined for each τ such that $1 \leqslant \tau \leqslant T$, the *zero coupon bond with maturity* τ is the security whose price at time τ is certain to be one. Its time-t price will be denoted Z_t^τ, so $Z^\tau = \{Z_t^\tau; 0 \leqslant t \leqslant \tau\}$ is an adapted process with $Z_\tau^\tau = 1$. The price of Z^τ is not defined for $t > \tau$.

The term structure model includes a zero coupon bond Z^τ for every τ satisfying $\tau = 1, \ldots, T$. Hence at each time t there is a collection $\{Z_t^{t+1}, Z_t^{t+2}, \ldots, Z_t^T\}$ of zero coupon bond prices. This collection is called the *term structure of zero coupon bond prices.*

The term structure model must be free of arbitrage opportunities, so there must exist a risk neutral probability measure Q under which the discounted prices of the zero coupon bonds are martingales. In other words, there must exist some probability measure Q with $Q(\omega) > 0$ for all $\omega \in \Omega$ such that, for every τ,

$$\textbf{(6.1)} \qquad Z_s^\tau = E_Q[B_s Z_t^\tau / B_t | \mathscr{F}_s], \qquad 0 \leqslant s \leqslant t \leqslant \tau$$

But $Z_\tau^\tau = 1$ and $B_t/B_s = (1 + r_{s+1}) \ldots (1 + r_t)$, so taking $t = \tau$ we see that zero coupon bonds must satisfy the important relationship

$$\textbf{(6.2)} \quad Z_s^\tau = E_Q[B_s/B_\tau | \mathscr{F}_s] = E_Q[1/\{(1 + r_{s+1}) \ldots (1 + r_\tau)\} | \mathscr{F}_s], \ 0 \leqslant s \leqslant \tau$$

given any risk neutral probability measure Q. Since $r_t > 0$, this implies, for each fixed s and ω, that $\tau \to Z_s^\tau(\omega)$ is a strictly decreasing function with $Z_s^{s+1}(\omega) < 1$. Note that taking $\tau = s + 1$ in (6.2) gives

$$\textbf{(6.3)} \qquad r_{s+1} + 1 = 1/Z_s^{s+1}, \qquad s = 0, 1, \ldots, T-1$$

With most kinds of securities market models it is customary to start with a probability space $(\Omega, \mathscr{F}, P, \mathbb{F})$ and then define the spot interest rate r and the risky securities with respect to the "real-world" probability measure P; this is because the future values of the risky securities are uncertain. Subsequently, a risk neutral probability measure Q is determined from these data. But with a term structure model, where future values of some key securities (i.e., the zero coupon bonds) are known with certainty, it is possible to take a different approach: first specify a probability space $(\Omega, \mathscr{F}, Q, \mathbb{F})$, letting Q be the risk neutral probability measure from the start, then specify the spot interest rate r, giving its probabilistic behavior with respect to Q, and finally use (6.2) to provide the specification of the zero coupon bond prices. Thus it is possible to construct a perfectly satisfactory term structure model without worrying about the probabilistic behavior of the spot interest rate and the zero coupon bond prices under a real-world probability measure.

Although this is a common approach, the reader should keep in mind that it is not immediately obvious whether there will exist a subjective probability measure P under which r and the zero coupon bond prices behave in a realistic or desirable manner. And it may be difficult to determine such a P, even if one is known to exist. This approach is illustrated in example 6.2, after first illustrating an alternative approach.

Example 6.1 With $T = 3$ and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_6\}$ suppose the time-1 partition is $\mathscr{P}_1 = \{\omega_1, \omega_2\} \cup \{\omega_3, \omega_4\} \cup \{\omega_5, \omega_6\}$ and the time-2 and time-3 partitions are $\mathscr{P}_2 = \mathscr{P}_3 = \{\omega_1\} \cup \ldots \cup \{\omega_6\}$. The "real-world" probability measure P can be specified, but its values are not important. For the spot interest rate, let $r_1 = 0.06$,

$$r_2(\omega) = \begin{cases} 0.09, & \omega = \omega_1, \omega_2 \\ 0.06, & \omega = \omega_3, \omega_4 \\ 0.03, & \omega = \omega_5, \omega_6 \end{cases}$$

and

$$r_3(\omega) = \begin{cases} 0.10, & \omega = \omega_1 \\ 0.08, & \omega = \omega_2 \\ 0.07, & \omega = \omega_3 \\ 0.05, & \omega = \omega_4 \\ 0.04, & \omega = \omega_5 \\ 0.02, & \omega = \omega_6 \end{cases}$$

Next is the specification of the zero coupon bond price processes Z^1, Z^2, and Z^3. For this you cannot take arbitrary stochastic processes with $Z_t^\tau < 1$ for $t < \tau$ and with $Z_\tau^\tau = 1$; you must carefully respect equation (6.1) in order to preclude the introduction of arbitrage opportunities. Indeed, by equation (6.3) there is no flexibility in the choice of $Z_{\tau-1}^\tau$ for $\tau = 1, 2$, and 3; these values are presented in table 6.1.

Next, consider the specification of $Z_1^3(\omega_1) = Z_1^3(\omega_2)$. Taking $t = s + 1$ in (6.1) we have

$$\textbf{(6.4)} \qquad Z_s^\tau = E_Q[(1 + r_{s+1})^{-1} Z_{s+1}^\tau | \mathscr{F}_s] = (1 + r_{s+1})^{-1} E_Q[Z_{s+1}^\tau | \mathscr{F}_s]$$

Now we have flexibility in the choice of Z_s^τ for $s < \tau - 1$, coming from the freedom to choose the conditional, risk neutral probabilities. In particular, with two possibilities for the value of $Z_2^3(\omega)$ given the information at time 1, $Z_1^3(\omega)$ can be chosen to be any value such that $(1 + r_2)Z_1^3(\omega)$ is strictly between the maximum and minimum values that the zero coupon price can become next period. For example, since $Z_2^3(\omega_1) = (1.1)^{-1} = 0.9091$ and $Z_2^3(\omega_2) = (1.08)^{-1} = 0.9259$, it suffices for $Z_1^3(\omega_1)$ and $Z_1^3(\omega_2)$ to satisfy the constraints

$$0.9091 < 1.09 Z_1^3(\omega_1) = 1.09 Z_1^3(\omega_2) < 0.9259$$

We shall take $Z_1^3(\omega_1) = Z_1^3(\omega_2) = 0.840$. In a similar fashion, we must have

$$Z_2^3(\omega_3) = 0.9346 < 1.06 Z_1^3(\omega_3) = 1.06 Z_1^3(\omega_4) < Z_2^3(\omega_4) = 0.9524$$

and

$$Z_2^3(\omega_5) = 0.9615 < 1.03 Z_1^3(\omega_5) = 1.03 Z_1^3(\omega_6) < Z_2^3(\omega_6) = 0.9804$$

so we shall take $Z_1^3(\omega_3) = Z_1^3(\omega_4) = 0.890$ and $Z_1^3(\omega_5) = Z_1^3(\omega_6) = 0.940$.

Turning to the specification of Z_0^2 and Z_0^3, we see that the situation is somewhat more complicated because each zero coupon bond can become one of three instead of just two distinct values. Equation (6.4) must be satisfied by both Z^2 and Z^3:

$$1.06 Z_0^2 = 0.9174p + 0.9434q + 0.9709(1 - p - q)$$
$$1.06 Z_0^3 = 0.84p + 0.89q + 0.94(1 - p - q)$$

where $p = Q(\omega_1) + Q(\omega_2)$ and $q = Q(\omega_3) + Q(\omega_4)$. These equations coupled with the constraints $p > 0$, $q > 0$, and $p + q < 1$ give rise to three constraints on the pair (Z_0^2, Z_0^3). We shall simply take $p = q = 0.3$, thereby giving $Z_0^2 = 0.893$ and $Z_0^3 = 0.844$.

This completes the specification of the zero coupon bond processes and thus of the model; the details are presented in the table 6.1. Notice that Z^3 is not an increasing process.

Table 6.1 Data for example 6.1

ω	Z_0^1	Z_0^2	Z_1^2	Z_0^3	Z_1^3	Z_2^3	$Q(\omega)$
ω_1	0.9434	0.893	0.9174	0.844	0.84	0.9091	0.1839
ω_2	0.9434	0.893	0.9174	0.844	0.84	0.9259	0.1161
ω_3	0.9434	0.893	0.9434	0.844	0.89	0.9346	0.1517
ω_4	0.9434	0.893	0.9434	0.844	0.89	0.9524	0.1483
ω_5	0.9434	0.893	0.9709	0.844	0.94	0.9615	0.2582
ω_6	0.9434	0.893	0.9709	0.844	0.94	0.9804	0.1418

It remains to compute a risk neutral probability measure Q. This will be done by first computing the conditional risk neutral probabilities, using equation (6.4). We already have these for the first period: $Q(\omega_1) + Q(\omega_2) = Q(\omega_3) + Q(\omega_4) = 0.3$ and $Q(\omega_5) + Q(\omega_6) = 0.4$. For the second period, equation (6.4) with $s = 1$, $\tau = 3$, and $\omega = \omega_1$ or ω_2 gives $1.09(0.84) = \hat{q}(0.9091) + (1 - \hat{q})\ (0.9259)$, so $\hat{q} = Q(Z_2^3 = 0.9091 | Z_1^3 = 0.84) = 0.6131$. In a similar fashion, equation (6.4) gives $Q(Z_2^3 = 0.9346 | Z_1^3 = 0.89) = 0.5056$ and $Q(Z_2^3 = 0.9615 | Z_1^3 = 0.94) = 0.6455$. Taking appropriate products of these conditional probabilities yields the risk neutral probability measure, which is displayed in table 6.1. Note this is the unique risk neutral probability measure associated with this specification of the zero coupon bonds, so this model is complete.

Example 6.2 With $T = 3$, the sample space Ω, the filtration $\mathbb{F}$, and the spot interest rate r all the same as in example 6.1, we can jump directly to the specification of the risk neutral probability measure Q. Any strictly positive probability measure will do. We shall illustrate this by simply taking $Q(\omega_1) = \ldots = Q(\omega_6) = 1/6$.

This completes the specification of the model (unless you also want the real-world probability measure P). It remains to derive the zero coupon bond price processes by using equation (6.2) or (6.4). The values of $Z_{\tau-1}^\tau$ for $\tau = 1, 2$, and 3 will, of course, be the same as in example 6.1. Equation (6.4) with $s = 1$, $\tau = 3$, and $\omega = \omega_1$ or ω_2 is

$$Z_1^3(\omega_1) = Z_1^3(\omega_2) = \frac{1}{2}\frac{0.9091}{1.09} + \frac{1}{2}\frac{0.9259}{1.09} = 0.8417$$

Similarly, $Z_1^3(\omega_3) = Z_1^3(\omega_4) = 0.8901$ and $Z_1^3(\omega_5) = Z_1^3(\omega_6) = 0.9427$. Equation (6.4) with $s = 0$ and $\tau = 3$ is

$$Z_0^3 = \frac{1}{3}\frac{0.8417}{1.06} + \frac{1}{3}\frac{0.8901}{1.06} + \frac{1}{3}\frac{0.9427}{1.06} = 0.8410$$

Finally, equation (6.4) with $s = 0$ and $\tau = 2$ is

$$Z_0^2 = \frac{1}{3}\frac{0.9174}{1.06} + \frac{1}{3}\frac{0.9434}{1.06} + \frac{1}{3}\frac{0.9709}{1.06} = 0.8905$$

The details for this model are summarized as follows:

ω	Z_0^1	Z_0^2	Z_1^2	Z_0^3	Z_1^3	Z_2^3	$Q(\omega)$
ω_1	0.9434	0.8905	0.9174	0.8410	0.8417	0.9091	0.1667
ω_2	0.9434	0.8905	0.9174	0.8410	0.8417	0.9259	0.1667
ω_3	0.9434	0.8905	0.9434	0.8410	0.8901	0.9346	0.1667
ω_4	0.9434	0.8905	0.9434	0.8410	0.8901	0.9524	0.1667
ω_5	0.9434	0.8905	0.9709	0.8410	0.9427	0.9615	0.1667
ω_6	0.9434	0.8905	0.9709	0.8410	0.9427	0.9804	0.1667

Notice that here, in contrast with example 6.1, all the zero coupon bond price processes are increasing.

The approaches used to specify the models in examples 6.1 and 6.2 are easy to implement, but there is no direct control over all the prices and interest rates that are observable at time $t = 0$. For practical purposes, it is often important to have a model where these time $t = 0$ values are equal to prescribed quantities. This is feasible to achieve, simply by starting with these values and then using equations (6.1) to (6.4) to introduce arbitrage-free future values of the interest rates and zero coupon bond prices. The risk neutral probability measure is introduced simultaneously, as will now be illustrated.

Example 6.3 With $T = 3$, the sample space Ω, and the filtration $\mathbb{F}$ all the same as in examples 6.1 and 6.2, suppose the values $r_1 = 0.06$, $Z_0^1 = 0.9434$, $Z_0^2 = 0.89$, and $Z_0^3 = 0.84$ are observed at time $t = 0$. To introduce future values of r and of the prices of these zero coupon bonds, it suffices to use equations (6.3) and (6.4), moving forward one period at a time; the risk neutral conditional probabilities will be introduced simultaneously.

Starting with the arbitrary values $Z_1^2(\omega_1) = 0.91$, $Z_1^2(\omega_3) = 0.94$, $Z_1^2(\omega_5) = 0.97$, and $Q(Z_1^2 = 0.91) = 0.3$, it follows from equation (6.4) with $s = 0$ and $\tau = 2$ that $Q(Z_1^2 = 0.94) = 0.2867$ and $Q(Z_1^2 = 0.97) = 0.4133$. Equation (6.3) thus implies that r_2 takes the values 0.0989, 0.0638, and 0.0309 in states ω_1, ω_3 and ω_5, respectively.

Turning next to the specification of Z_1^3, we see that we can choose arbitrary values for two of the three possible values, say $Z_1^3(\omega_3) = 0.89$ and $Z_1^3(\omega_5) = 0.94$. This choice implies $Z_1^3(\omega_1) = 0.8223$, since equation (6.4) must be satisfied with $s = 0$ and $\tau = 3$.

The next step is to choose the values of Z_2^3, making sure equation (6.4) with $s = 1$ and $\tau = 3$ is satisfied for suitable values of the conditional, risk neutral probabilities. For instance, with $Z_2^3(\omega_1) = 0.90$ and $Z_2^3(\omega_2) = 0.92$, equation (6.4) gives $Q(Z_2^3 = 0.90 | Z_1^3 = 0.8223) = 0.8187$. Similarly, the choice of $Z_2^3(\omega_3) = 0.94$, $Z_2^3(\omega_4) = 0.95$, $Z_2^3(\omega_5) = 0.96$, and $Z_2^3(\omega_6) = 0.98$ implies $Q(Z_2^3 = 0.94 | Z_1^3 = 0.89) = 0.32$ and $Q(Z_2^3 = 0.96 | Z_1^3 = 0.94) = 0.55$. Equation (6.3) stipulates the values of r_3. This model is summarized as follows:

ω	Z_1^2	Z_1^3	Z_2^3	r_2	r_3	Q
ω_1	0.91	0.8223	0.90	0.0989	0.1111	0.2456
ω_2	0.91	0.8223	0.92	0.0989	0.0870	0.0544
ω_3	0.94	0.89	0.94	0.0638	0.0638	0.0917
ω_4	0.94	0.89	0.95	0.0638	0.0526	0.1950
ω_5	0.97	0.94	0.96	0.0309	0.0417	0.2273
ω_6	0.97	0.94	0.98	0.0309	0.0204	0.1860

It remains to compute the risk neutral probability measure Q. This follows immediately from the conditional probabilities which have already been computed; the values of $Q(\omega)$ are shown above.

We now turn to a new topic, the *yield to maturity*. This is an adapted stochastic process, denoted $Y^\tau = \{Y_t^\tau;\ t = 0, \ldots, \tau - 1\}$, that is uniquely associated with each zero coupon bond. The value Y_t^τ is defined to be the one-period interest rate such that a sum of money equal to the current price of the zero coupon bond, namely Z_t^τ, will become exactly $Z_\tau^\tau = 1$ at time τ when invested and compounded at this constant rate. In other words,

$$Z_t^\tau (1 + Y_t^\tau)^{\tau - t} = 1, \qquad 0 \leqslant t < \tau \leqslant T$$

which is the same as

$$\textbf{(6.5)} \qquad Z_t^\tau = (1 + Y_t^\tau)^{t - \tau}, \qquad 0 \leqslant t < \tau \leqslant T$$

as well as

$$\textbf{(6.6)} \qquad Y_t^\tau = [Z_t^\tau]^{1/t - \tau} - 1, \qquad 0 \leqslant t < \tau \leqslant T$$

Notice that $Y_t^{t+1} = r_{t+1}$, the current spot interest rate.

At each time t there is a collection $\{Y_t^{t+1}, \ldots, Y_t^T\}$ of yields that is called the *term structure of interest rates* or the *yield curve*. In view of equations (6.5) and (6.6), knowledge of the term structure of interest rates is equivalent to knowledge of the term structure of zero coupon bond prices $\{Z_t^{t+1}, \ldots, Z_t^T\}$.

Example 6.1 (continued) Using equation (6.6), we quickly derive the following yields

ω	Y_0^1	Y_0^2	Y_0^3	Y_1^2	Y_1^3	Y_2^3
ω_1	0.06	0.0582	0.0582	0.09	0.0911	0.10
ω_2	0.06	0.0582	0.0582	0.09	0.0911	0.08
ω_3	0.06	0.0582	0.0582	0.06	0.0600	0.07
ω_4	0.06	0.0582	0.0582	0.06	0.0600	0.05
ω_5	0.06	0.0582	0.0582	0.03	0.0314	0.04
ω_6	0.06	0.0582	0.0582	0.03	0.0314	0.02

Notice the various kinds of term structures. The time $t = 0$ term structure is decreasing with respect to maturity, two of the time $t = 1$ term structures are increasing, and the third time $t = 1$ term structure is constant.

Another important concept pertaining to term structure models is that of *forward interest rates.* Suppose it is time s, and consider the forward price O_s of the τ-maturity zero coupon bond, delivered at time t, where $s \leqslant t \leqslant \tau$. In view of principle (4.22) and equation (6.2), this must be

$$O_s = \frac{Z_s^\tau}{E_Q[B_s/B_t | \mathscr{F}_s]} = \frac{Z_s^\tau}{Z_s^t}, \qquad 0 \leqslant s \leqslant t \leqslant \tau \leqslant T$$

This equation makes economic sense, since there are no arbitrage opportunities. In one case you buy the τ-maturity discount bond at time s and hold it until time τ. In the other case you make at time s a forward contract to take delivery of this same bond at time t and then hold it until it matures at time τ, financing the time t payment O_s by investing at time s exactly $O_s Z_s^t$ dollars in the t-maturity discount bond. In both cases you are sure to have \$1 at time τ, so the time s expenditures must be the same by the law of one price.

For the special case where $\tau = t + 1$,

$$O_s = \frac{Z_s^{t+1}}{Z_s^t}, \quad 0 \leqslant s \leqslant t \leqslant T \tag{6.7}$$

must be the time-s forward price of a zero coupon bond that is delivered at time t and matures one period later. The yield, denoted $f(s, t)$, corresponding to forward price (6.7) must be

$$f(s, t) = \frac{Z_s^t}{Z_s^{t+1}} - 1 \tag{6.8}$$

because Z_s^{t+1}/Z_s^t dollars invested at time t at the interest rate $f(s,t)$ will become one dollar at time $t + 1$. Note that $f(s, t) > 0$, because, as pointed out earlier, $t \to Z_s^t$ is a strictly decreasing function. Since the yield $f(s,t)$ is associated with a single period, it will be called the *forward spot interest rate* or, simply, the *forward interest rate.*

Taking $t = s$ in (6.8), we see from (6.3) that

$$f(s, s) = r_{s+1}, \qquad 0 \leqslant s < T$$

This is logically consistent, because if delivery occurs right away, then the forward and spot interest rates coincide. We also see from (6.8) that $f(s, t)$ is an $\mathscr{F}_s$– measurable random variable for each $t \geqslant s$. Hence, for each fixed t, $s \to f(s, t)$ is an adapted stochastic process.

The collection $\{f(s, s), \ldots, f(s, T-1)\}$ is called the time-s *term structure of forward interest rates.* In view of equation (6.8), knowledge of the term structure of zero coupon bond prices gives you the term structure of forward interest rates. The converse is true, because $Z_s^s = 1$ and so you can use (6.8) to work out the remaining zero coupon bond prices in a recursive manner. Hence all three kinds of term structures are equivalent.

A useful relationship between the forward rates and the price of a discount bond is obtained from the identity

$$Z_s^\tau = \prod_{t=s+1}^{\tau} [Z_s^t / Z_s^{t-1}]$$

Substituting (6.8) yields

$$Z_s^\tau = \prod_{t=s+1}^{\tau} [1 + f(s, t-1)]^{-1} \tag{6.9}$$

Compare this with equation (6.2).

Example 6.1 (continued) Using equation (6.8), we quickly derive the following forward interest rates.

ω	$f(0,0)$	$f(0,1)$	$f(0,2)$	$f(1,1)$	$f(1,2)$	$f(2,2)$
ω_1	0.06	0.0564	0.0581	0.09	0.0921	0.10
ω_2	0.06	0.0564	0.0581	0.09	0.0921	0.08
ω_3	0.06	0.0564	0.0581	0.06	0.0600	0.07
ω_4	0.06	0.0564	0.0581	0.06	0.0600	0.05
ω_5	0.06	0.0564	0.0581	0.03	0.0329	0.04
ω_6	0.06	0.0564	0.0581	0.03	0.0329	0.02

This is consistent with equation (6.9), because, for example,

$$Z_0^2 = \frac{1}{(1.06)(1.0564)} = 0.893$$

Exercise 6.1 For a term structure model with $T = 5$, suppose the time $t = 0$ term structure is as indicated below. Derive the other two kinds of term structures.

(a) The term structure of zero coupon bond prices is $Z_0^1 = 0.96$, $Z_0^2 = 0.915$, $Z_0^3 = 0.88$, $Z_0^4 = 0.837$, and $Z_0^5 = 0.80$.
(b) The term structure of yields is $Y_0^1 = 4\%$, $Y_0^2 = 5\%$, $Y_0^3 = 6\%$, $Y_0^4 = 7\%$, and $Y_0^5 = 6.5\%$.
(c) The term structure of forward interest rates is $f(0,0) = 4\%$, $f(0,1) = 6\%, f(0,2) = 5.5\%, f(0,3) = 5\%$, and $f(0,4) = 5\%$.

Exercise 6.2 For the same model as in example 6.2 except that $Q(\omega_i) = i/21$ for $i = 1, \ldots, 6$, compute all the zero coupon bond, yield, and forward rate processes. (Hint: show $Z_0^3 = 0.8604$)

6.2 Lattice, Markov Chain Models

Given an arbitrary filtration $\mathbb{F}$, it is usually difficult to specify the spot interest rate process r and other elements of the model so that the various interest rate and price processes will be realistic under a subjective probability measure P and, moreover, their time-0 values will be consistent with an arbitrary time-0 term structure. It is desirable to develop a class of simple models where these objectives are met.

An obvious candidate that comes to mind is an interest rate version of the binomial model studied in section 3.5. The idea would be to model the spot interest rate r just like the risky security in the binomial model, selecting a value q for the risk neutral conditional probability, and then the zero coupon bond prices would follow from equation (6.2). However, due to the fact this model involves only four parameters (namely, r_1, u, d, and q), there is no reason to expect the computed zero coupon bond prices to match an arbitrary term structure. This kind of model is too simple.

But all is not lost with the binomial model approach. A modest generalization will do the job. One retains the same submodel of the information structure as for the binomial model, namely, as portrayed in figure 3.4. But the spot interest rate process r will be generalized to the extent of being a Markov chain, along the lines of the risky security of example 3.6 in section 3.6. To be precise, let X denote a Markov chain with initial value $X_0 = 0$, with state space $E = \{0, 1, \ldots, T\}$, and with transition probabilities satisfying

$$P\{X_{t+1} = j | X_t = n\} > 0, \quad j = n+1 \text{ or } j = n,$$
$$P\{X_{t+1} = j | X_t = n\} = 0, \quad \text{otherwise}$$

for $t = 0, 1, \ldots, T-1$. Hence X_t can be thought of as the cumulative number of heads after t tosses of a coin, but unlike the binomial process N_t studied in sections 3.5 and 3.6, the various coin flips are not necessarily independent or identically distributed. Instead, the probability that coin flip number $t+1$ is a head depends, in general, on both t and the current value X_t.

Figure 6.1 shows a network with the nodes representing the states which X can possibly reach at each time t starting with X_0. The branches correspond to positive transition probabilities. Needless to say, this picture is essentially the same as the lattice in figure 3.5 for the binomial model.

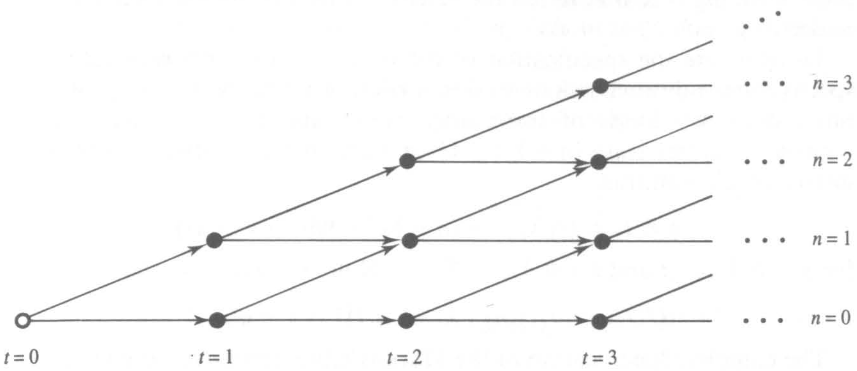

Figure 6.1. State space for the Markov chain X

The Markov chain X is not stationary, because the transition probabilities can vary with respect to t. However, upon defining a new stochastic process by setting $\hat{X}_t = (X_t, t)$, it is apparent that $\hat{X}$ is also a Markov chain that is, in fact, stationary. Indeed, the nodes in figure 6.1 correspond to its possible states, each of the form (n, t), and the branches correspond to its positive Markov transition probabilities. This "stationary" perspective will be more convenient for the term structure model that is being developed.

Now suppose there are functions $\rho_t : \{0, 1, \ldots, t\} \to (0, \infty)$ for $t = 0, 1, \ldots, T-1$. Usually, each such function is strictly increasing on its domain, but this is not crucial. We can define the spot interest rate process r by setting

$$r_{t+1}(\omega) = \rho_t(X_t(\omega)), \; t = 0, 1 \ldots, T-1$$

The Markov chain X can be interpreted as an exogenous "factor" with the property that knowledge of it implies knowledge of the spot interest rate r. Knowing that $X_t = n$ at time t, say, you not only know that $r_{t+1} = \rho_t(n)$, but you also know that at time $t+1$ the spot interest rate r_{t+2} will be either $\rho_{t+1}(n)$ or $\rho_{t+1}(n+1)$. In particular, if each function ρ_t is strictly increasing on its domain, then knowledge of r_{t+1} is equivalent to knowledge of X_t, and the spot interest rate process is itself a Markov chain. This is the usual case.

Of course, explicit knowledge of the functions ρ_t is not necessary; they were introduced primarily as a device to link this term structure model with the discussion in section 3.6 about Markov chains, especially example 3.5. To develop the term structure model one can proceed directly from the state space representation as displayed in figure 6.1 to the specification of the spot interest rate process r_t, simply by specifying its value at each node of the lattice. Let $r_{t+1}(n)$ denote the value associated with state (n,t) for $n = 0, 1 \ldots, t$ and $t = 0, 1, \ldots T-1$. Even if $r_{t+1}(n_1) = r_{t+1}(n_2)$, say, the

understanding is that at time t the agents in this securities market know the underlying state $\hat{X}_t = (n, t)$.

To complete the specification of the term structure model it suffices to specify the conditional risk neutral transition probabilities for the process $\hat{X}$. Since only two kinds of transitions from state $\hat{X}_t = (n, t)$ are possible, namely, to either state $(n+1, t+1)$ or state $(n, t+1)$, it is convenient to introduce the notation

$$q(n, t) = Q\{\hat{X}_{t+1} = (n+1, t+1) | \hat{X}_t = (n, t)\}$$

for $n = 0, 1, \ldots, t$ and $t = 0, 1, \ldots, T-1$. Note, for example, that

$$Q\{r_{t+2} = r_{t+2}(n) \mid \hat{X}_t = (n, t)\} = 1 - q(n, t)$$

The complete specification of the Markov chain term structure model can now be neatly displayed with the lattice diagram shown in figure 6.2. With T time periods, there are $1 + 2 + \ldots + T = T(T+1)/2$ nodes in this lattice, not counting the terminal nodes at time T. Thus there is a total of $T(T+1)$ parameters to be specified in this model, one value of r and one value of q for each node.

What about the prices of the zero coupon bonds? In view of equation (6.2) and the Markov property, we have that

$$\begin{aligned} Z_t^\tau &= E_Q[1/\{(1+r_{t+1})\ldots(1+r_\tau)\}| \mathscr{F}_t] \\ &= E_Q[1/\{(1+r_{t+1})\ldots(1+r_\tau)\}|X_t] \end{aligned}$$

In words, the price of a zero coupon bond depends on the state of the underlying factor X_t, but otherwise it is independent of the history of prices and interest rates. As with the spot interest rate, the value of Z_t^τ can be expressed as some function of X_t. Equivalently, the process Z^τ is fully

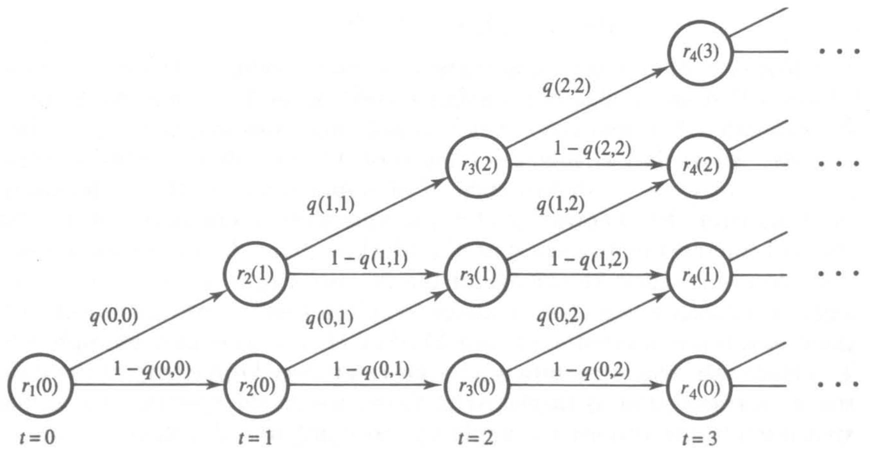

Figure 6.2 The Markov chain term structure model

specified by knowing its value at each node of the lattice in figure 6.2. We shall take this latter approach and let $Z_t^\tau(n)$ denote the value of Z_t^τ at the node corresponding to $X_t = n$.

It is useful to have a formula for $Z_t^\tau(n)$ in terms of the spot interest rate and conditional risk neutral probabilities. We can derive one from equation (6.4), which now can be rewritten as

$$\textbf{(6.10)} \qquad Z_t^\tau(n) = \frac{1}{1 + r_{t+1}(n)}\left[q(n, t) Z_{t+1}^\tau(n+1) + [1 - q(n, t)] Z_{t+1}^\tau(n)\right]$$

Denote

$$\delta(n, t, 1) = \frac{q(n, t)}{1 + r_{t+1}(n)} \quad \text{and} \quad \delta\,(n, t, 0) = \frac{1 - q(n, t)}{1 + r_{t+1}(n)}$$

so that (6.10) can be simplified to become

$$Z_t^\tau(n) = \sum_{i=0}^{1} \delta\,(n, t, i)\, Z_{t+1}^\tau(n+i)$$

Since $Z_\tau^\tau = 1$, a backwards induction argument on t can be used with this equation to show for $t = 0, \ldots, \tau$ and $n = 0, \ldots, t$ that

$$\textbf{(6.11)} \qquad \begin{aligned} Z_t^\tau(n) = &\sum_{i_1=0}^{1} \delta(n, t, i_1) \sum_{i_2=0}^{1} \delta\,(n + i_1, t+1, i_2) \ldots \\ &\ldots \sum_{i_{\tau-t}=0}^{1} \delta\,(n + i_1 + \ldots + i_{\tau-t-1}, \tau - 1, i_{\tau-t}) \end{aligned}$$

In particular, the time-0 prices of the zero coupon bonds are given by

$$\textbf{(6.12)} \qquad \begin{aligned} Z_0^\tau = &\sum_{i_1=0}^{1} \delta\,(0, 0, i_1) \sum_{i_2=0}^{1} \delta\,(i_1, 1, i_2) \ldots \\ &\ldots \sum_{i_\tau=0}^{1} \delta\,(i_1 + \ldots + i_{\tau-1}, \tau - 1, i_\tau) \end{aligned}$$

As mentioned at the beginning of this section, it is desirable for the term structure model to be consistent with the observed, time-0 term structure of zero coupon bond prices. In principle, equation (6.12) can be used to choose the parameters of the model in a consistent manner. But there are two problems with this. First, equation (6.12) is difficult to work with for the purpose of solving for parameters on the right hand side. Secondly, our general Markov term structure model entails $T(T+1)$ parameters, whereas the observed time-0 term structure together with (6.12) give only T equations in the $T(T+1)$ unknowns. We still have a great many degrees of freedom in the choice of the parameters.

This matter will be resolved by making further simplifications in the model. The idea is to reduce the effective number of parameters by making

assumptions about the spot interest rate and/or the conditional risk neutral probabilities. The aim is to obtain a system of equations which is feasible to solve for unique values of the parameters.

The usual simplification is to make some assumption about the "volatilities," which are taken to be either the ratios $r_{t+1}(n+1)/r_{t+1}(n)$ or the ratios $[1+r_{t+1}(n+1)]/[1+r_{t+1}(n)]$. These two approaches are examined in the following two examples.

Example 6.4 Assume the conditional risk neutral probabilities are independent of n, that is,

$$q(n,t) = q(t), \quad 0 \leqslant n \leqslant t < T$$

These T numbers $q(0), \dots, q(T-1)$, and thus all the conditional risk neutral probabilities, are specified. In addition, T numbers $c(0), \dots, c(T-1)$ are also specified, and it is required that (the first equality here is just (6.3))

$$\textbf{(6.13)} \qquad \frac{Z_t^{t+1}(n+1)}{Z_t^{t+1}(n)} = \frac{1+r_{t+1}(n)}{1+r_{t+1}(n+1)} = c(t), \quad 0 \leqslant n \leqslant t < T$$

The number $c(t)$ can be interpreted as a measure of the volatility of the time-t spot interest rate, a measure which is independent of n. Using (6.13) recursively gives

$$\textbf{(6.14)} \quad Z_{\tau-1}^{\tau}(n+i) = Z_{\tau-1}^{\tau}(n)c^i(\tau-1), \quad 0 \leqslant n \leqslant n+i \leqslant \tau-1 < T$$

In particular,

$$\textbf{(6.15)} \qquad Z_{\tau-1}^{\tau}(i) = Z_{\tau-1}^{\tau}(0)c^i(\tau-1), \quad 0 \leqslant i \leqslant \tau-1 < T$$

Since $Z_{\tau-1}^{\tau}(n) = 1/[1+r_\tau(n)]$, this says that knowing $r_t(0)$ you can deduce $r_t(n)$ for all $n \geqslant 1$. Hence this model will be fully specified if we can choose the T numbers $r_0(0), \dots, r_{T-1}(0)$ (equivalently, $Z_0^1(0)$, $Z_1^2(0), \dots, Z_{T-1}^T(0)$) to be consistent with the T observed zero coupon bond prices $Z_0^1, \dots, Z_0^T$.

To show how to do this, use will be made of the following equation:

$$\textbf{(6.16)} \qquad Z_t^\tau(n) = \prod_{j=t}^{\tau-1} g(j,\tau-1)Z_j^{j+1}(n), \qquad n = 0,1,\dots,t$$

where $g(s,s) \equiv 1$ and

$$g(j,s) \quad \equiv 1 - q(j) + q(j)c(j+1)\dots c(s), \quad j = 0,1,\dots,s-1$$

Equation (6.16) can be verified with a backwards induction argument on t, beginning with $t = \tau-1$, for which it is clearly true. Assuming (6.16) is true for $t+1$, equation (6.10) gives

$$\begin{aligned}
Z_t^\tau(n) &= Z_t^{t+1}(n)\Big[q(t)Z_{t+1}^\tau(n+1) + [1-q(t)]Z_{t+1}^\tau(n)\Big] \\
&= Z_t^{t+1}(n)\Big[q(t)\prod_{j=t+1}^{\tau-1} g(j,\tau-1)\,Z_j^{j+1}(n+1) \\
&\qquad + \{1-q(t)\}\prod_{k=t+1}^{\tau-1} g(k,\tau-1)\,Z_k^{k+1}(n)\Big] \\
&= Z_t^{t+1}(n)\Big[q(t)\prod_{j=t+1}^{\tau-1} g(j,\tau-1)\,Z_j^{j+1}(n)c(j) \\
&\qquad + \{1-q(t)\}\prod_{k=t+1}^{\tau-1} g(k,\tau-1)Z_k^{k+1}(n)\Big]
\end{aligned}$$

where the last equality uses (6.14). Hence

$$\begin{aligned}
Z_t^\tau(n) &= Z_t^{t+1}(n)\Big[q(t)\prod_{j=t+1}^{\tau-1} c(j) + 1 - q(t)\Big]\prod_{j=t+1}^{\tau-1} g(j,\tau-1)Z_j^{j+1}(n) \\
&= Z_t^{t+1}(n)g(t,\tau-1)\prod_{j=t+1}^{\tau-1} g(j,\tau-1)\,Z_j^{j+1}(n)
\end{aligned}$$

and equation (6.16) is verified.

It follows from (6.16) that

$$\frac{Z_0^{t+1}(0)}{Z_0^t(0)} = \frac{\prod_{j=0}^{t} g(j,t)Z_j^{j+1}(0)}{\prod_{j=0}^{t-1} g(j,t-1)Z_j^{j+1}(0)} = \frac{Z_t^{t+1}(0)\prod_{j=0}^{t-1} g(j,t)}{\prod_{j=0}^{t-2} g(j,t-1)}$$

where the second equality uses $g(t,t) = g(t-1,t-1) = 1$. Hence this and (6.15) give the key result for $0 \leqslant n \leqslant t < T$:

$$\textbf{(6.17)} \qquad Z_t^{t+1}(n) = \frac{Z_0^{t+1}(0)}{Z_0^t(0)}[c(t)]^n \frac{\prod_{j=0}^{t-2} g(j,t-1)}{\prod_{j=0}^{t-1} g(j,t)}$$

The right hand side of (6.17) is completely known, coming from either the specified numbers $q(t)$ and $c(t)$ or the time-0 term structure of zero coupon bond prices. Hence $r_{t+1}(n) = 1/Z_t^{t+1}(n) - 1$ is known for all n and t.

The numbers $c(t)$ are normally in the vicinity of 1, which means the spot interest rate does not change too much from one period to the next. Indeed, these numbers must be selected with care in order to keep the interest rates from becoming negative or unreasonably large. If, for instance, $c(t) = 1$, then by (6.14) $Z_t^{t+1}(n)$ is independent of n. This also implies $g(j,t) = g(j,t-1)$, so (6.17) becomes $Z_t^{t+1}(n) = Z_0^{t+1}(0)/Z_0^t(0)$, that is, one period zero coupon bond prices at time t do not depend on the state n but only on the initial term structure.

This model can be generalized to allow the risk neutral conditional probabilities $q(t)$ to depend on the state n. In the opposite direction, an important special case is obtained by taking $q(t) = q$ and $c(t) = 1/k$ for all t, where q and k are specified positive scalars with $q < 1$. Then for $j < s$ the function g becomes

$$g(j,s) = 1 - q + qk^{j-s} \equiv \frac{1}{h(s-j)}$$

where the new function h is defined in an obvious manner. In this case (6.17) becomes simply $Z_t^{t+1}(n) = Z_0^{t+1}(0)k^{-n}h(t)/Z_0^t(0)$.

Example 6.5 Assume $q(n,t) = 0.5$ for all n and t, so

$$\delta(n,t,1) = \delta(n,t,0) = \frac{0.5}{1 + r_{t+1}(n)} = 0.5Z_t^{t+1}(n)$$

Equation (6.12) simplifies to become

$$Z_0^\tau = \left(\frac{1}{2}\right)^{\tau-1} Z_0^1(0)\sum_{i_1=0}^{1} Z_1^2(i_1) \sum_{i_2=0}^{1} Z_2^3(i_1+i_2)\ldots \ldots \sum_{i_{\tau-1}=0}^{1} Z_{\tau-1}^\tau(i_1+\ldots+i_{\tau-1}) \tag{6.18}$$

It remains to specify the $T(T+1)/2$ values of the spot interest rate. They will be chosen to be consistent with the T time-0 zero coupon bond prices as well as $T(T-1)/2$ specified spot rate volatilities, giving a total of $T(T+1)/2$ constraints.

The spot rate volatility is defined to be

$$\sigma_t(n) \equiv \frac{1}{2}\ln\left(\frac{r_{t+2}(n+1)}{r_{t+2}(n)}\right), \quad 0 \leqslant n \leqslant t \leqslant T-1 \tag{6.19}$$

Thus if you know r_{t+1} or $Z_t^{t+1}(n)$ for one value of n, you know both for all $n = 0, \ldots, t$, just like example 6.4. For instance, $r_{t+2}(n+1) = r_{t+2}(0)\exp\{2[\sigma_t(0) + \ldots + \sigma_t(n)]\}$. Using this idea, equation (6.18), and the time 0 term structure, you can work out all the values of the spot interest rate by moving forward in time in a recursive manner. Beginning with $\tau = 2$ in (6.18) and $t = 0$ in (6.19), you solve for the two unknowns $r_2(0)$ and $r_2(1)$. In general, knowing $Z_{t-1}^t(n)$ and $r_t(n)$ for all $0 \leqslant n < t < \tau$, you use $\tau - 1$ versions of (6.19) with $t = \tau - 2$ together with (6.18) to solve for the τ unknowns $r_\tau(0), \ldots, r_\tau(\tau - 1)$.

An alternative perspective is to view the spot interest rate process r as being governed by the stochastic difference equation

$$\Delta r_{t+1} = \mu(t, r_t) + \sigma(t, r_t)N_t, \tag{6.20}$$

where $\mu : \mathbb{R}^2 \to \mathbb{R}$ and $\sigma : \mathbb{R}^2 \to \mathbb{R}_+$ are specified functions, $\{N_t\}$ is a sequence of independent and identically distributed random variables with $Q(N_t = 1) = Q(N_t = -1) = 0.5$, and, as usual, $\Delta r_{t+1} = r_{t+1} - r_t$. This stochastic difference equation formulation is appealing, because under the risk neutral probability measure Q the conditional expected change in the spot rate is $\mu(t, r_t)$ and the conditional variance of the change in the spot rate is $\sigma^2(t, r_t)$.

Given a spot interest rate process r defined on a binary lattice, it is easy to come up with consistent functions for μ and σ. For example, knowing $r_t(n)$, $r_{t+1}(n)$, and $r_{t+1}(n+1)$, the stochastic difference equation gives

$$r_{t+1}(n+1) - r_t(n) = \mu(t, r_t(n)) + \sigma(t, r_t(n))$$

and

$$r_{t+1}(n) - r_t(n) = \mu(t, r_t(n)) - \sigma(t, r_t(n))$$

in which case one must have

$$\mu(t, r_t(n)) = \frac{r_{t+1}(n+1) + r_{t+1}(n) - 2r_t(n)}{2}$$

and

$$\sigma(t, r_t(n)) = \frac{r_{t+1}(n+1) - r_{t+1}(n)}{2}$$

More useful is to start with a stochastic difference equation for r and then derive a binomial lattice formulation. For example, suppose

$$\Delta r_{t+1} = \phi(t) - ar_t + \sigma(t)N_t, \tag{6.21}$$

where a is a scalar satisfying $0 < a < 1$ and ϕ and σ are positive functions with domain $\mathbb{R}$. This formulation is appealing, because it says the interest rate tends to decrease if $r_t > \phi(t)/a$, otherwise the interest rate tends to increase. However, the function σ must satisfy a certain constraint in order for this process to be consistent with a binomial lattice formulation. In particular, an *up-down* move (that is, $N_t = 1$ and $N_{t+1} = -1$) must arrive at the same value for the spot interest rate as a *down-up* move (that is, $N_t = -1$ and $N_{t+1} = 1$). In other words, we must have

$$\begin{aligned} r_{t+2} &= \phi(t+1) + (1-a)\phi(t) + (1-a)^2 r_t + (1-a)\sigma(t) - \sigma(t+1) \\ &= \phi(t+1) + (1-a)\phi(t) + (1-a)^2 r_t - (1-a)\sigma(t) + \sigma(t+1) \end{aligned}$$

for all t. A necessary and sufficient (assuming $\sigma > 0$) condition for this to be true is that $\sigma(t+1) = (1-a)\sigma(t) = (1-a)^t\sigma(1)$. Hence the adaptation of this stochastic difference equation formulation to a binomial lattice leads to a severe restriction on the "volatility" functions σ.

This restriction on the volatility function can be ameliorated by working with a trinomial lattice. The idea here is that from one period to the next the spot interest rate moves to one of three possible values: "up," "middle," or "down." Moreover, various paths recombine: up-down = middle-middle = down-up, up-middle = middle-up, and middle-down = down-middle. Thus

there will be three nodes after one period, five nodes after two periods, and, in general, $2t+1$ nodes after t periods.

The stochastic difference equation for r has the same form as (6.20), only now the random variables $\{N_t\}$ are independent but satisfy $Q(N_t = 1) = Q(N_t = -1) = q(t)$ and $Q(N_t = 0) = 1 - 2q(t)$, where $0 < q(t) < 1/2$. Hence (see exercise 6.7) the conditional mean and variance of the change in the spot interest rate are $\mu(t, r_t)$ and $2q(t)\sigma^2(t, r_t)$, respectively.

Now consider again the special case (6.21), only with $\{N_t\}$ as in the preceding paragraph. This formulation is compatible with a trinomial lattice if and only if (see exercise 6.8) $\sigma(t+1) = (1-a)\sigma(t) = (1-a)^t\sigma(1)$, the same condition as before. In this case, the conditional variance of the change of the spot interest rate is $2q(t)(1-a)^{2t-2}\sigma(1)$. Hence flexibility in the choice of the probabilities $q(t)$ give the model builder some room to match the conditional variances with desired values. For instance, these variances can be made constant by taking $q(t) = 0.5(1-a)^{2-2t}$, provided these probabilities satisfy $0 < q(t) < 1/2$.

Exercise 6.3 Show for example 6.4 that if $m_t^{-1} \leqslant [c(t)]^t \leqslant m_t$ for some number $m_t > 1$, then

$$\frac{Z_0^{t+1}(0)}{m_t Z_0^t(0)} \leqslant Z_t^{t+1}(n) \leqslant \frac{m_t Z_0^{t+1}(0)}{Z_0^t(0)}$$

Use this fact to show how the c's can be selected to ensure the spot interest rate remains between specified upper and lower bounds.

Exercise 6.4 Show for the special case of example 6.4 with $q(t) = q$ and $c(t) = 1/k$ that starting with arbitrary values of q, k, and the initial term structure, it is possible to obtain negative interest rates.

Exercise 6.5 For the model in example 6.4, show that

(a) $Z_t^\tau(n) = Z_t^\tau(0)\prod_{j=t}^{\tau-1} c^n(j)$, $n \leqslant t < \tau$

(b) $Z_t^\tau(n) = [Z_0^\tau(0)/Z_0^t(0)]\left[\prod_{j=t}^{\tau-1} c^n(j)\right]\left[\prod_{j=0}^{t-1}\frac{g(j,t-1)}{g(j,\tau-1)}\right]$, $n \leqslant t < \tau$

(c) $f(t,\tau,n) = [Z_0^\tau(0)/Z_0^{\tau+1}(0)]\left[c^{-n}(\tau)\right]\left[\prod_{j=0}^{t-1}\frac{g(j,\tau)}{g(j,\tau-1)}\right] - 1$, $n \leqslant t < \tau$

where $f(t,\tau,n)$ denotes the forward spot rate corresponding to state n at time t.

Exercise 6.6 Example 6.5 can be generalized to allow for arbitrary conditional probabilities with $0 < q(n,t) < 1$, in which case the volatilities are defined by

$$\sigma_t(n) = \ln\{r_{t+2}(n+1)/r_{t+2}(n)\}\sqrt{q(n,t)[1-q(n,t)]}$$

For an arbitrary term structure of discount bond prices with $Z_0^1 > Z_0^2 > \ldots > Z_0^T$ and for arbitrary volatilities with $\sigma_t(n) > 0$, show by induction that there exists a unique and strictly positive spot interest rate

process r such that the model is consistent with these quantities and free of arbitrage.

Exercise 6.7 Suppose the spot interest rate r is governed by the stochastic difference equation (6.20) with $Q(N_t = 1) = Q(N_t = -1) = q(t)$ and $Q(N_t = 0) = 1 - 2q(t)$, where $0 < q(t) < 1/2$. Verify that the conditional mean and variance of the change in the spot interest rate are $\mu(t, r_t)$ and $2q(t)\sigma^2(t, r_t)$, respectively.

Exercise 6.8 Suppose the spot interest rate r is governed by the stochastic difference equation (6.21) with $Q(N_t = 1) = Q(N_t = -1) = q(t)$ and $Q(N_t = 0) = 1 - 2q(t)$, where $0 < q(t) < 1/2$. Assuming $\sigma(t) > 0$, verify that this formulation is compatible with a trinomial lattice if and only if $\sigma(t+1) = (1-a)\sigma(t) = (1-a)^t\sigma(1)$.

Exercise 6.9 Suppose the spot interest rate is governed by stochastic difference equation (6.20) with the volatility function of the form $\sigma(t,r) = \sqrt{r}\,\sigma(t)$ for some positive function σ on $\mathbb{R}$. Under what restrictions on μ and σ will this formulation be compatible with:

(a) a binomial lattice?
(b) a trinomial lattice?

6.3 Yield Curve Models

In the preceding two sections, the construction of term structure models emphasized the role of the spot interest rate process $r = \{r_t; t \geqslant 1\}$. The idea was to first specify the probability space and the filtration, to then specify the process r (usually it is taken to be a Markov chain), and to finally use equations like (6.2) to specify the zero coupon bond processes. This is an easy approach to implement, providing a variety of arbitrage-free term structure models. However, this approach has a disadvantage: it is difficult to model the behavior of yields and zero coupon bond prices at different maturities. For example, the model builder may be concerned with the volatility of the zero coupon bond that matures in period 10. Or the model builder may be particularly interested in the spread between short and long term interest rates. Features like these cannot be modeled explicitly with the spot interest rate approach.

An alternative approach, called the *yield curve* or *whole yield* approach, is to build the model by regarding the whole term structure as the state of a stochastic process. After specifying the probability space and the filtration submodel of the information flow, you directly specify how the whole term structure evolves in time. This can be done by working with either the term structure of yields, the term structure of zero coupon bond prices, or the term structure of forward spot rates. Whatever the choice, after one term structure process is specified, the other two as well as the spot interest rate processes follow from the equations in section 6.1. Moreover,

the term structure process is usually taken to be a Markov chain, and one can work with either the real-world or the risk neutral probability measure.

A disadvantage of the yield curve approach is that worrying about arbitrage opportunities can make its development and implementation more difficult than with the spot interest rate approach. Combining equations (6.3) and (6.4) gives

$$Z_s^\tau = Z_s^{s+1} E_Q[Z_{s+1}^\tau | \mathscr{F}_s] \;,\; 0 \leqslant s < \tau \leqslant T \tag{6.22}$$

For the absence of arbitrage opportunities, it is necessary and sufficient that there exists a probability measure Q (the risk neutral probability measure) such that this equation is satisfied for all indicated s and τ. But as illustrated in the following example, an arbitrary selection of the term structure process may not satisfy (6.22) for any probability measure.

Example 6.6 Suppose $T = 3$, $K = 8$, and the information submodel is a binomial tree (but not a lattice). The time 0 zero coupon bond prices are taken to be $Z_0^1 = 0.95$, $Z_0^2 = 0.90$, and $Z_0^3 = 0.85$. In the case of an "up" move, the time 1 discount bond prices are taken to be $Z_1^2 = 0.94$ and $Z_1^3 = 0.89$, whereas in the case of a "down" move the time 1 discount bond prices are taken to be $Z_1^2 = 0.96$ and $Z_1^3 = 0.91$. Let q denote the risk neutral conditional probability of an "up" move between times 0 and 1. Equation (6.22) for $\tau = 2$ implies $q = 0.6316$, whereas equation (6.22) for $\tau = 3$ implies $q = 0.7632$. But there is no probability measure Q under which (6.22) is satisfied for both $\tau = 2$ and $\tau = 3$. Hence there must exist an arbitrage opportunity.

To produce an arbitrage opportunity, let H_i denote the units purchased at time 0 of the zero coupon bond that matures after i periods, $i = 1, 2$, and 3. The initial cost of this portfolio is $V_0 = 0.95H_1 + 0.9\, H_2 + 0.85H_3$; for an arbitrage opportunity we want $V_0 = 0$, implying, for instance, $H_2 = -(19/18)\, H_1 - (17/18)\, H_3$. After an up move the time 1 value of the portfolio will be

$$V_1 = H_1 + 0.94\, H_2 + 0.89\, H_3 = 0.0078\, H_1 + 0.0022\, H_3$$

whereas after a down move the value will be

$$V_1 = H_1 + 0.96\, H_2 + 0.91\, H_3 = -\,0.0133\, H_1 + 0.0033\, H_3$$

Thus any trading strategy which makes the right hand sides of both values of V_1 strictly positive will be an arbitrage opportunity. For instance, one could take $H_1 = 0$ and $H_3 = 18$, in which case $H_2 = -17$.

The problem illustrated by example 6.6 was that the model was overspecified, making it impossible for (6.22) to be satisfied. To circumvent this difficulty, the trick is to specify just enough of the risk neutral probability measure Q and the term structure process values so that (6.22) (or one of its equivalents) can be used to uniquely specify the balance of the

model. This will be illustrated by working with a binary tree submodel of the information, where the risk neutral conditional probability of an "up" move from a node will always be denoted by q (the actual value can possibly vary from one node to another). The objective here is to explain how to achieve a satisfactory specification of the term structure process of zero coupon bond prices. Later, an alternative, more efficient approach will be presented, where the constructed process is the term structure of forward spot rates.

Suppose the term structure $\{Z_s^{s+1}, Z_s^{s+2}, \ldots, Z_s^T\}$ of zero coupon bond prices at a time s node of the information tree has been specified. This can be any collection of $T - s$ numbers satisfying $1 > Z_s^{s+1} > Z_s^{s+2} > \ldots > Z_s^T > 0$. The objective, to be carried out in two steps, is to specify the time $s+1$ values of these zero coupon bond prices at the two "downstream" nodes. This two-step process can be replicated at the other nodes in the information tree, thereby completing the specification of the term structure process.

The first step is to set the time $s+1$ zero coupon bond prices in the event of an up move. These will be $T - s - 1$ arbitrary numbers denoted $Z_{s+1}^{s+2}(u)\, Z_{s+1}^{s+3}(u) \ldots Z_{s+1}^{T}(u)$ and satisfying

$$1 > Z_{s+1}^{s+2}(u) > Z_{s+1}^{s+3}(u) > \ldots > Z_{s+1}^{T}(u) > 0 \tag{6.23}$$

It is here that the model builder can incorporate desirable features, such as the manner in which the term structure shifts from one time period to the next.

The second and final step is to use equation (6.22) to specify the $T - s - 1$ zero coupon bond prices, denoted $Z_{s+1}^{s+2}(d)$, $Z_{s+1}^{s+3}(d), \ldots, Z_{s+1}^{T}(d)$, in the event of a "down" move. Solving for $Z_{s+1}^{\tau}(d)$, (6.22) gives

$$Z_{s+1}^{\tau}(d) = \frac{Z_s^\tau / Z_s^{s+1} - qZ_{s+1}^{\tau}(u)}{1-q} \;,\; \tau = s+2, \ldots, T \tag{6.24}$$

Hence this completes the specification of the term structure process at the two downstream nodes, except for one possible problem: there is no guarantee that the term structure of zero coupon bond prices will be "legitimate," satisfying the inequalities

$$1 > Z_{s+1}^{s+2}(d) > Z_{s+1}^{s+3}(d) > \ldots > Z_{s+1}^{T}(d) > 0 \tag{6.25}$$

In view of (6.24), it is not sufficient for the zero prices after an up move to merely satisfy the inequalities in (6.23); they must satisfy some additional constraints.

Equation (6.24) says that $Z_{s+1}^{T}(d) > 0$ if and only if

$$Z_{s+1}^{T}(u) < \frac{Z_s^T}{qZ_s^{s+1}} \tag{6.26}$$

This is easy to satisfy, since the right hand side is sure to be strictly positive.

Equation (6.24) says that $1 > Z_{s+1}^{s+2}(d)$ if and only if

$$(6.27) \qquad Z_{s+1}^{s+2}(u) > \frac{Z_s^{s+2}}{qZ_s^{s+1}} - \frac{1-q}{q}$$

This also is easy to satisfy, since an easy computation shows that the right hand side is sure to be strictly less than one.

Finally, the remaining inequalities in (6.25) are

$$Z_{s+1}^{\tau}(d) > Z_{s+1}^{\tau+1}(d),\ \tau = s+2, \ldots, T-1$$

Equation (6.24) says that these inequalities are satisfied if and only if

$$(6.28) \qquad Z_{s+1}^{\tau}(u) - Z_{s+1}^{\tau+1}(u) < \frac{Z_s^{\tau} - Z_s^{\tau+1}}{qZ_s^{s+1}},\ \tau = s+2, \ldots, T-1$$

Given arbitrary values of $q, Z_s^{s+1}, \ldots,$ and Z_s^T with $q \in (0,1)$ and $1 > Z_s^{s+1} > \ldots > Z_s^T > 0$, it is possible to choose $Z_{s+1}^{s+2}(u), \ldots, Z_{s+1}^T(u)$ so as to satisfy (6.28) as well as (6.23), (6.26), and (6.27) (see exercise 6.10). Similarly, given arbitrary values of $Z_s^{s+1}, \ldots,$ and Z_s^T with $1 > Z_s^{s+1} > \ldots > Z_s^T > 0$ and arbitrary values of $Z_{s+1}^{s+2}(u), \ldots, Z_{s+1}^T(u)$ satisfying (6.23), it is possible to choose q close enough to 0 so that (6.28) will be satisfied. Hence with a little bit of flexibility, it is always possible to specify legitimate term structures of zero coupon bond prices at both downstream nodes. Moreover, since equation (6.22) was utilized, the absence of arbitrage opportunities is ensured.

Example 6.7 Suppose $Z_s^{s+i} = \delta^i$ and $Z_{s+1}^{s+1+i}(u) = \theta^i$ for $i = 1, 2, \ldots$ and two numbers $\delta, \theta \in (0,1)$. The numbers δ and T are fixed; there is some flexibility in the choice of q and θ. Inequality (6.26) is the same as

$$(6.29) \qquad q\,\theta^{T-s-1} < \delta^{T-s-1}$$

whereas inequality (6.27) is the same as

$$(6.30) \qquad 1 - q + q\,\theta > \delta$$

Moreover, the inequalities in (6.28) are the same as

$$(6.31) \qquad q(1-\theta)\theta^{\tau-s-1} < (1-\delta)\delta^{\tau-s-1},\ \tau = s+2, \ldots, T-1$$

Now suppose $\theta = \delta + \varepsilon$ for some $\varepsilon > 0$. Inequality (6.29) will be true provided $q < (\delta/\theta)^{T-s-1}$, so reasonable values of q can be accommodated if ε is small enough. Inequality (6.30) is automatically true, because $1 - q + q\theta > 1 - q + q\delta > (1-q)\delta + q\delta = \delta$. Inequality (6.31) is true for all $\tau \leqslant T-1$ if and only if it is true for $\tau = T-1$, so our final requirement on q and ε is

$$q < \frac{(1-\delta)}{(1-\delta-\varepsilon)}\left(\frac{\delta}{\delta+\varepsilon}\right)^{T-s-2}$$

a requirement that is usually easy to achieve with reasonable values of q and ε.

Constructing the term structures of zero coupon bond prices at each node of a binary tree can be tricky, because you need to worry about constraints of the form (6.25). It is often easier to work with term structures of forward spot rates, because instead of constraints like (6.25) you only need to make sure the forward spot rates at each node are positive (of course, in both cases you must also worry about the no-arbitrage conditions).

To see how to construct a forward spot rate yield curve model, suppose a term structure $\{f(s,s), \ldots, f(s, T-1)\}$ of (non-negative) forward spot rates has been specified at a time s node of the binary information tree. Let q be the conditional risk neutral probability of an "up" move to the time $(s+1)$ node where the term structure $\{f_u(s+1, s+1), \ldots, f_u(s+1, T-1)\}$ has also been specified. The objective is to use the no-arbitrage conditions to produce the term structure $\{f_d(s+1, s+1), \ldots, f_d(s+1, T-1)\}$ at the time $(s+1)$ "down" node.

Using equation (6.9) to substitute for Z in equation (6.22), one obtains the no-arbitrage condition

$$(6.32) \qquad \prod_{t=s+2}^{\tau} [1 + f(s, t-1)]^{-1}$$
$$= E_Q\left[\prod_{t=s+2}^{\tau} [1 + f(s+1, t-1)]^{-1} | \mathscr{F}_s\right],\ \tau = s+2, \ldots, T$$

We can use these $T-s-1$ equations to solve for the $T-s-1$ variables $f_d(s+1, s+1), \ldots, f_d(s+1, T-1)$. Denote $g(s,t) \equiv [1 + f(s,t)]^{-1}$, $g_u(s,t) \equiv [1 + f_u(s,t)]^{-1}$, and $g_d(s,t) \equiv [1 + f_d(s,t)]^{-1}$. Then (6.32) can be rewritten as

$$\prod_{t=s+2}^{\tau} g(s, t-1) = q \prod_{t=s+2}^{\tau} g_u(s+1, t-1) + (1-q) \prod_{t=s+2}^{\tau} g_d(s+1, t-1)$$

in which case

$$(6.33) \qquad g_d(s+1, \tau-1)$$
$$= \frac{\prod_{t=s+2}^{\tau} g(s, t-1) - q \prod_{t=s+2}^{\tau-1} g_u(s+1, t-1)}{(1-q) \prod_{t=s+2}^{\tau} g_d(s+1, t-1)},\ \tau = s+2, \ldots, T$$

Hence the g_d's can be computed recursively, starting with $\tau = s+2$ (the denominator is just $1-q$ in this case), then $\tau = s+3$, and so forth. Given arbitrary values of q, the f's, and the f_u's, there is no guarantee that all the f_d's will turn out positive. However, given arbitrary values of the f's and some flexibility in the choice of q and the f_u's, it is always possible to obtain non-negative f_d's.

Exercise 6.10 Verify the statement made just after (6.28): given arbitrary values of q, $Z_s^{s+1}, \ldots,$ and Z_s^T with $q \in (0,1)$ and $1 > Z_s^{s+1} > \ldots > Z_s^T > 0$, it is possible to choose $Z_{s+1}^{s+2}(u), \ldots, Z_{s+1}^T$ so as to satisfy (6.28) as well as (6.23), (6.26), and (6.27).

Exercise 6.11 In example 6.7, suppose $s = 0$, $\delta = 0.95$, and $T = 8$. What values of $\theta > 0.95$ are admissible if $q = 0.5$? What values of q are admissible if $\theta = 0.96$? Compute $Z_1^\tau(d)$, $\tau = 2, \ldots, T$, for the case where $q = 0.5$ and $\theta = 0.96$.

Exercise 6.12 Suppose $s = 0$, $q = 0.5$, $T = 5$, $f(0,0) = 0.05$, $f(0,1) = 0.06$, $f(0,2) = 0.07$, $f(0,3) = 0.06$, and $f(0,4) = 0.05$. Moreover, suppose $f_u(1,t) = f(0,t) + 0.01$ for all $t \geqslant 1$. Use (6.33) to verify that $f_d(1,1) = 0.0502$, $f_d(1,2) = 0.0603$, $f_d(1,3) = 0.0505$, and $f_d(1,4) = 0.0407$.

Exercise 6.13 Suppose positive values of $f(s,s), \ldots, f(s,T-1)$ are fixed. For the cases listed below, explain whether it is always possible to produce positive values of $f_d(s+1,s+1), \ldots, f_d(s+1,T-1)$ satisfying (6.33). If yes, give a proof. If no, give a counter-example.

(a) $q \in (0,1)$ and positive values of $f_u(s+1,s+1), \ldots, f_u(s+1,T-1)$ are all fixed, arbitrary numbers.
(b) $q \in (0,1)$ is a fixed number, but there is flexibility in the choice of the positive numbers $f_u(s+1,s+1), \ldots, f_u(s+1,T-1)$.
(c) The positive numbers $f_u(s+1,s+1), \ldots, f_u(s+1,T-1)$ are fixed, but there is flexibility in the choice of $q \in (0,1)$.

6.4 Forward Risk Adjusted Probability Measures

In order to facilitate computations for interest rate derivatives, this section will introduce and describe the properties of a new probability measure. In particular, a new and quite useful formula will be developed for the price of a contingent claim. First, however, are some preliminary results from probability theory.

Throughout this section fix the time $\tau \leqslant T$ and consider a strictly positive random variable $M_\tau \in \mathscr{F}_\tau$ satisfying $E_Q[M_\tau] = 1$. For the time being, Q here is an arbitrary probability measure, although shortly we shall take Q to be the risk neutral probability measure.

Define a new probability measure, denoted P_τ, by setting

$$P_\tau(\omega) \equiv M_\tau(\omega)Q(\omega), \quad \text{all } \omega \in \Omega$$

Note that P_τ is indeed a legitimate probability measure, because $P_\tau(\omega) > 0$ for all ω and the assumption $E_Q[M_\tau] = 1$ implies $P_\tau(\Omega) = 1$. Let E_τ denote the expectation operator corresponding to P_τ.

Next, define a martingale $M = \{M_t; t = 0, \ldots, \tau\}$ by setting

$$M_t \equiv E_Q[M_\tau|\mathscr{F}_t], \quad t = 0, 1, \ldots, \tau.$$

Notice that $M_0 = E_Q[M_\tau] = 1$ and that M is a martingale with respect to Q but not necessarily with respect to P_τ. This martingale plays a role in the following technical result, which relates conditional expectation under the two probability measures.

(6.34) If X is a random variable, then $E_\tau[M_t X|\mathscr{F}_t] = E_Q[M_\tau X|\mathscr{F}_t]$ for $t = 0, 1, \ldots, \tau$.

Clearly (6.34) is true for $X \in \mathscr{F}_t$, for then the equation is simply $M_t X = X E_Q[M_\tau|\mathscr{F}_t]$. It is also easily seen to be true for $t = 0$, for then $M_0 = 1$ and we have

$$E_\tau[X] = \sum_\omega X(\omega)P_\tau(\omega) = \sum_\omega X(\omega)M_\tau(\omega)Q(\omega) = E_Q[M_\tau X]$$

For the general case, it suffices to consider arbitrary $A \in \mathscr{P}_t$, an event in the partition of Ω corresponding to $\mathscr{F}_t$, and show that

$$E_\tau[M_t X|A] = E_Q[M_\tau X|A] \tag{6.35}$$

For the left hand side we have

$$E_\tau[M_t X|A] = \frac{\sum_{\omega\in A} X(\omega)M_t(\omega)P_\tau(\omega)}{\sum_{\omega\in A} P_\tau(\omega)} = \frac{\sum_{\omega\in A} X(\omega)M_t(\omega)M_\tau(\omega)Q(\omega)}{\sum_{\omega\in A} M_\tau(\omega)Q(\omega)}$$

But M_t is constant on A and given by

$$M_t(\omega) = E_Q[M_t|A] = \sum_{\hat\omega\in A} M_\tau(\hat\omega)Q(\hat\omega)/Q(A), \quad \text{all } \omega \in A$$

so substituting this in the expression for $E_\tau[M_t X|A]$ yields

$$E_\tau[M_t X|A] = \frac{\sum_{\omega\in A} X(\omega)M_\tau(\omega)Q(\omega)}{Q(A)} = E_Q[M_\tau X|A]$$

This completes the verification of (6.35) and thus (6.34).

We are now in a position to explain a fundamental relationship in probability theory:

(6.36) The stochastic process $YM = \{Y_t M_t; t = 0, \ldots, \tau\}$ is a martingale under Q if and only if the stochastic process $Y = \{Y_t; t = 0, \ldots, \tau\}$ is a martingale under P_τ.

To see this, note that YM is a martingale under Q if and only if $Y_t M_t = E_Q[Y_\tau M_\tau|\mathscr{F}_t]$ for all t. Now using (6.34) with $X = Y_\tau$, we see the latter is true if and only if $Y_t M_t = E_\tau[M_t Y_\tau|\mathscr{F}_t]$ for all t. But $E_\tau[M_t Y_\tau|\mathscr{F}_t] = M_t E_\tau[Y_\tau|\mathscr{F}_t]$, so the latter is the same as $Y_t = E_\tau[Y_\tau|\mathscr{F}_t]$ for all t, that is, Y is a martingale under P_τ.

With these preliminaries out of the way, we are ready to return to our term structure model. Let the stochastic process $\pi = \{\pi_t : 0 \leqslant t \leqslant s\}$ represent the price of an asset such as a stock, a zero coupon bond, or a contingent claim, where $\tau \leqslant s \leqslant T$. Set $Y_t = \pi_t/Z_t^\tau$ and recall from principle (4.22) that Y_t

represents the time-t *forward price* for delivery of the asset at time τ. Using our standard notation for forward prices, we therefore will sometimes write O_t for $Y_t = \pi_t/Z_t^\tau$.

Next, let Q be the risk neutral probability measure and set $M_\tau = [B_\tau Z_0^\tau]^{-1}$. Note that $M_\tau(\omega) > 0$ and $E_Q[M_\tau] = (1/Z_0^\tau)E_Q[1/B_\tau] = 1$ (because $Z_0^\tau = E_Q[1/B_\tau]$). Hence we can proceed as above and define the Q-martingale

$$M_t = E_Q[M_\tau|\mathcal{F}_t] = \frac{1}{Z_0^\tau}E_Q[1/B_\tau|\mathcal{F}_t] = \frac{Z_t^\tau}{Z_0^\tau B_t}$$

where the last equality follows from the risk neutral formula (6.1) for Z_t^τ. We also define the *forward risk adjusted probability measure* (also called the τ *forward probability measure*):

$$P_\tau(\omega) = M_\tau(\omega)Q(\omega) = \frac{Q(\omega)}{Z_0^\tau B_\tau(\omega)}$$

Now observe that $Y_t M_t = O_t M_t = (\pi_t/Z_t^\tau)(Z_t^\tau/[Z_0^\tau B_t]) = \pi_t/[Z_0^\tau B_t]$, so the process YM represents the discounted price of the asset divided by the constant Z_0^τ. This is a martingale under the risk neutral probability measure Q, so by (6.36) we have the following important result:

(6.37) Under the forward risk adjusted probability measure P_τ, the time-t forward price O_t for delivery of an asset at time τ is a martingale.

Principle (6.37) is important because it leads to a new and useful formula for the price of a derivative security. Let $\pi_\tau \in \mathcal{F}_\tau$ be the time $-\ \tau$ price of a security; for instance, π_τ is the time $-\ \tau$ payoff of a contingent claim. But the time $-\ \tau$ forward price O_τ for time $-\ \tau$ delivery of this security is $O_\tau = \pi_\tau$, so (6.37) implies

$$O_t = \pi_t/Z_t^\tau = E_\tau[O_\tau|\mathcal{F}_t] = E_\tau[\pi_\tau|\mathcal{F}_t], \quad t \leqslant \tau$$

Multiplying through by Z_t^τ yields the following:

(6.38) If π_t is the time-t price of a security, then

$$\pi_t = Z_t^\tau E_\tau[\pi_\tau|\mathcal{F}_t], \quad t \leqslant \tau$$

The traditional risk neutral formula $\pi_t = B_t E_Q[\pi_\tau/B_\tau|\mathcal{F}_t]$ is convenient when the spot interest rate is constant or deterministic, for then the time$-\tau$ value B_τ of the bank account process is deterministic and this formula simplifies to $\pi_t = (B_t/B_\tau)E_Q[\pi_\tau|\mathcal{F}_t]$. Hence for derivative pricing all you need is the conditional distribution of π_τ under the risk neutral probability measure. But for interest rate models and other situations where the spot interest rate r is stochastic, the bank account value B_τ does not factor outside the conditional expectation, and so to apply the traditional formula you need the conditional *joint* distribution of (π_τ, B_τ) under the risk neutral probability measure. In practice, this may be difficult to obtain. On the other hand, even with stochastic interest rates, to apply the formula in (6.38), all you need is the conditional distribution of π_τ under the forward risk adjusted probability measure corresponding to time$-\tau$.

Example 6.1 (continued) For $\tau = 3$ we first compute $M_3(\omega) = [Z_0^3 B_3(\omega)]^{-1} = [0.844\{1 + r_1(\omega)\}\{1 + r_2(\omega)\}\{1 + r_3(\omega)\}]^{-1}$. Then using the risk neutral probabilities we compute the forward risk adjusted probability measure $P_3(\omega) = M_3(\omega)Q(\omega)$. These numbers are given below, along with the earlier values of the martingale M, which are computed using $M_t = E_Q[M_3|\mathcal{F}_t]$.

ω	B_3	M_0	M_1	M_2	M_3	P_3
ω_1	1.2709	1.0	0.939	0.9323	0.9323	0.1714
ω_2	1.2478	1.0	0.939	0.9495	0.9495	0.1102
ω_3	1.2023	1.0	0.995	0.9855	0.9855	0.1495
ω_4	1.1798	1.0	0.995	1.0043	1.0043	0.1489
ω_5	1.1355	1.0	1.051	1.0434	1.0434	0.2694
ω_6	1.1136	1.0	1.051	1.0640	1.0640	0.1509

To illustrate the application of formula (6.38), consider the contingent claim X with time-3 payoff $X(\omega_i) = i, i = 1, \ldots, 6$. The computation of π_2 is easy, because $\pi_2 = Z_2^3 E_3[X|\mathcal{F}_2] = Z_2^3 X$. Thus $\pi_2(\omega_6) = Z_2^3(\omega_6)X(\omega_6) = 0.9804(6) = 5.8824$, and similarly $\pi_2(\omega_i) = 0.9091, 1.8518, 2.8038, 3.8096$ and 4.8075 for $i = 1, \ldots, 5$, respectively.

For $\pi_1 = Z_1^3 E_3[X|\mathcal{F}_1]$ we need the conditional probability distribution, but this is easily computed from P_3. For example,

$$\begin{aligned} E_3[X|\{\omega_1,\omega_2\}] &= 1P_3(\omega{=}\omega_1|\{\omega_1,\omega_2\}) + 2P_3(\omega{=}\omega_2|\{\omega_1,\omega_2\}) \\ &= \frac{P_3(\omega_1)}{P_3(\omega_1)+P_3(\omega_2)} + 2\frac{P_3(\omega_2)}{P_3(\omega_1)+P_3(\omega_2)} \\ &= 0.6087 + 2(0.3913) = 1.3913 \end{aligned}$$

so $\pi_1(\omega_1) = \pi_1(\omega_2) = Z_1^3(\omega_1)E_3[X|\{\omega_1,\omega_2\}] = 0.84(1.3913) = 1.1687$. Similarly, $\pi_1(\omega_3) = \pi_1(\omega_4) = 3.1141$ and $\pi_1(\omega_5) = \pi_1(\omega_6) = 5.0375$.

To complete the computation of the process π, we have $\pi_0 = Z_0^3 E_3[X|\mathcal{F}_0] = 0.844E_3[X]$, so we simply use the probability measure P_3 itself and compute $\pi_0 = 3.113$.

Finally, using $O_t = \pi_t/Z_t^3$ we immediately have the forward price process: $O_0 = 3.6884$, $O_1(\omega_1) = O_1(\omega_2) = 1.3913$, $O_1(\omega_3) = O_1(\omega_4) = 3.4990$, $O_1(\omega_5) = O_1(\omega_6) = 5.3590$, and $O_2(\omega_i) = O_3(\omega_i) = i$ for $i = 1, \ldots, 6$. Using $o_t = E_3[o_3|\mathcal{F}_t]$ and the conditional probabilities corresponding to P_3, it is straightforward to verify that the process O is indeed a martingale under P_3.

Example 6.2 (continued) In order to prepare for the computation of some derivatives having known time-2 payoffs, we are interested in the corresponding forward risk adjusted probability measure P_2. With $M_2 = [Z_0^2 B_2]^{-1} = [0.8905(1 + r_1)(1 + r_2)]^{-1}$ and $Q(\omega) = 1/6$ for all ω, the quantities of interest are:

ω	B_2	M_2	P_2
ω_1	1.1554	0.9719	0.1620
ω_2	1.1554	0.9719	0.1620
ω_3	1.1236	0.9994	0.1666
ω_4	1.1236	0.9994	0.1666
ω_5	1.0918	1.0285	0.1714
ω_6	1.0918	1.0285	0.1714

Computing forward risk adjusted probability measures for lattice-type interest rate models is not much easier than it is for general models. This is because although the spot interest rate process r is a path-independent Markov chain, with each node of the lattice corresponding to a value of the spot rate at a point in time, the bank account process B is not path-independent. The different paths leading up to a node will correspond to different sequences of spot rates and thus to different bank account values. In general, for a binomial lattice, B_t will take one of 2^{t-1} values, whereas there will only be $t+1$ nodes for time t. Hence when computing the forward risk adjusted probability measure, the lattice loses most of its simplifying advantages. The computations required for a lattice model are not much easier than for a general model having about the same number of states ω. In particular, there are no simple formulas for the forward risk adjusted probability measures associated with the binomial lattice models of section 6.2.

Exercise 6.14 For example 6.1, verify with detailed calculations that:

(a) M is a martingale under Q with $E_Q[M_3] = 1$.
(b) The computed price process π satisfies (6.38) for all t and ω.
(c) The computed price process π satisfies the traditional risk neutral valuation formula for all t and ω.
(d) The computed forward price process O is a martingale under P_3.

Exercise 6.15 For example 6.1, suppose a derivative security has time-3 price $\pi_3(\omega_i) = \max\{i-3, 0\}, i = 1, \ldots, 6$. Compute the time-0 price π_0.

Exercise 6.16 For example 6.2, suppose there is an asset with time-2 prices $\pi_2(\omega_1) = \pi_2(\omega_2) = 7$, $\pi_2(\omega_3) = \pi_2(\omega_4) = 9$, and $\pi_2(\omega_5) = \pi_2(\omega_6) = -1$. Show that $\pi_0 = 4.3848$, and compute π_1.

Exercise 6.17 For example 6.2, specify the Q-martingale M and the forward risk adjusted probability measure P_3 corresponding to $\tau = 3$.

Exercise 6.18 With the process M as in (6.36), show that the process $X = \{X_t; t = 0, \ldots, \tau\}$ is a martingale under P_τ, where $X_t(\omega) = 1/M_t(\omega)$, all ω and t.

Exercise 6.19 If $X = 1_A$ is the time$-\tau$ payoff of a contingent claim, where $A \in \mathscr{F}_\tau$, then show its time-0 price is $\pi_0 = Z_0^\tau P_\tau(A)$.

6.5 Coupon Bonds and Bond Options

Consider a European call option on the zero coupon bond Z^s having time-τ payoff:

$$X = (Z_\tau^s - K)^+, \quad 0 \leqslant \tau < s \leqslant T$$

Our aim is to compute the time-t price π_t of X for $t < \tau$. Using (6.38), this is given by

$$\pi_t = Z_t^\tau E_\tau[(Z_\tau^s - K)^+ \mid \mathscr{F}_t], \quad t \leqslant \tau \tag{6.39}$$

where E_τ denotes expectation with respect to the forward risk adjusted probability measure P_τ. Hence we first need the conditional probability distribution of Z_τ^s under P_τ, and then we can compute the conditional expectation.

Alternatively, of course, we can use the customary risk neutral valuation formula:

$$\pi_t = B_t E_Q[(Z_\tau^s - K)^+ / B_\tau | \mathscr{F}_t], \quad t \leqslant \tau \tag{6.40}$$

The specific circumstances will dictate which approach will be the easiest from the computational standpoint.

Example 6.2 (continued) For a general model such as this, the forward risk adjusted approach is often the best, especially if one is planning to value more than one derivative having the same payoff time. Suppose $K = 0.95$, $t = 0$, $\tau = 2$, and $s = 3$. Then (6.39) gives

$$\begin{aligned}\pi_0 &= 0.8905 E_2[(Z_2^3 - 0.95)^+] \\ &= 0.8905\{0.1666(0.0024) + 0.1714(0.0115) + 0.1714(0.0304)\} \\ &= 0.0068\end{aligned}$$

Alternatively, using (6.40) we have

$$\begin{aligned}\pi_0 &= E_Q[(Z_2^3 - 0.95)^+ / B_2] \\ &= 0.1667(0.0024)/1.1236 + 0.1667(0.0115)/1.0918 \\ &\quad + 0.1667(0.0304)/1.0918 \\ &= 0.0068\end{aligned}$$

It was remarked at the end of section 6.4 that lattice models of interest rates offer no simple formulas for forward risk adjusted probability measures because the bank account process is path dependent. For the same reason, the same remark holds true in regard to the use of the risk neutral formula (6.40): there are no simple formulas for the price of contingent claims such as our call option on a bond when the underlying model is a lattice model as in section 6.2.

A coupon bond is just a linear combination of zero coupon bonds. To see this, suppose a coupon bond is scheduled to pay C_n dollars at time t_n for $n = 1, \ldots, N$, where $t < t_1 < \ldots < t_N$. By either the risk neutral valuation formula (6.1) or the forward risk adjusted formula (6.38), the time-t price of the contingent claim which pays C_n at time t_n is precisely $Z_t^{t_n} C_n$. Moreover, the price of the cash flow is just the sum of the price of each component, so with β_t denoting the time-t price of this bond, we have

$$\beta_t = \sum_{n=1}^{N} Z_t^{t_n} C_n$$

In words, the price of the coupon bond is equal to the expected discounted value of the cash flow, where expectation is with respect to the risk neutral probability measure.

Example 6.1 (continued) Suppose $C_1 = 7$, $t_1 = 2$, $C_2 = 107$, and $t_2 = 3$. Then $\beta_0 = 7Z_0^2 + 107Z_0^3 = 96.559$ and

$$\beta_1(\omega) = 7Z_1^2 + 107Z_1^3 = \begin{cases} 96.302, & \omega = \omega_1, \omega_2 \\ 101.834, & \omega = \omega_3, \omega_4 \\ 107.376, & \omega = \omega_5, \omega_6 \end{cases}$$

Now suppose we have a European call option on this coupon bond. The time-τ payoff, where $t < \tau < t_1$, is, as usual, $X = (\beta_\tau - K)^+$. For the time-t price, we can use (6.1), giving

$$\pi_t = B_t E_Q\Big[\Big(\sum_{n=1}^{N} Z_\tau^{t_n} C_n - K\Big)^+ / B_\tau | \mathscr{F}_t\Big] \tag{6.41}$$

Or we can use (6.38), giving

$$\pi_t = Z_t^\tau E_\tau\Big[\Big(\sum_{n=1}^{N} Z_\tau^{t_n} C_n - K\Big)^+ | \mathscr{F}_t\Big] \tag{6.42}$$

There are rarely any computational shortcuts with either the forward risk adjusted or the risk neutral approach.

Example 6.1 (continued) Suppose we have a call option on the coupon bond with exercise price $K = 100$ and exercise date $\tau = 1$. Equation (6.41) gives for the time-0 price

$$\begin{aligned} \pi_0 &= (0.1517 + 0.1483)(1.834/1.06) + (0.2582 + 0.1418)(7.376/1.06) \\ &= 3.3025 \end{aligned}$$

Exercise 6.20 Let c and p denote the time-0 prices of a European call and a put, respectively, both having the same exercise date τ, the same exercise price K, and the same underlying, namely, a zero coupon bond with maturity date $s > t$. Show that the following put-call parity relationship holds: $c - p = Z_0^s - KZ_0^\tau$. Compute p in the case of example 6.2 when $K = 0.95$, $\tau = 2$, and $s = 3$. Verify the put-call parity relationship for this particular put and the corresponding call.

Exercise 6.21 For example 6.1, compute the time-0 prices of the European put and call having exercise price $K = 0.95$, exercise date $\tau = 2$, and underlying Z^3. Verify the corresponding put-call parity relationship.

Exercise 6.22 For example 6.2, compute the time-0 and time-1 prices of the coupon bond which pays 8 dollars at time 2 and 108 dollars at time 3. Compute the time-0 prices of the European put and call on this bond, where the exercise price is $K = 102$ and the exercise date is $\tau = 1$.

6.6 Swaps and Swaptions

Swaps are agreements between two parties where the first pays a floating rate to the second, while the second pays a fixed rate to the first, with both payments being based on a common principal amount. The payments are made each period during an interval of time. The floating rate payment is based on the spot rate r, with the value actually used being either the one for the period just ended (the swap is *settled in arrears*) or the one for the period about to begin (the swap is *settled in advance*).

With ordinary swaps the initial floating rate payment is based on the spot rate when the agreement is made; this is true whether it is settled in arrears or in advance. There are also *forward start* swaps where the initial payments are based on a spot rate that is subsequent to the one which exists when the agreement is reached.

The value of a swap is just the expected present value of the net cash flow, so the value to one party is the negative of the value to the opposite party. With a *payer* swap the value is from the perspective of the party who pays the fixed rate and receives the floating rate. A *receiver* swap is the opposite.

This section will focus on a payer forward start swap on principal 1 settled each period in arrears, leaving other cases for the reader. The fixed interest rate is denoted K. With the initial floating rate payment based on r_τ, the party will pay K dollars and receive r_τ dollars at time τ. Similar payments will occur each period through time-s, so the time-t value of this payer forward start swap is

$$\pi_t = E_Q\left[\sum_{u=\tau}^{s} \frac{B_t}{B_u}(r_u - K)\Big|\mathscr{F}_t\right], \quad t < \tau \leqslant s \leqslant T \tag{6.43}$$

It turns out there is a simple and useful formula for π_t. Since $Z_{u-1}^u = (1 + r_u)^{-1}$, this value equals

$$\pi_t = E_Q\left[\sum_{u=\tau}^{s} \frac{B_t}{B_u}\left(\frac{1}{Z_{u-1}^u} - (1+K)\right)\bigg|\mathscr{F}_t\right]$$
$$= E_Q\left[\sum_{u=\tau}^{s}\left(\frac{B_t}{B_u Z_{u-1}^u}\right)\bigg|\mathscr{F}_t\right] - (1+K)E_Q\left[\sum_{u=\tau}^{s}\frac{B_t}{B_u}\bigg|\mathscr{F}_t\right]$$
$$= \sum_{u=\tau}^{s} E_Q\left[\frac{B_t}{B_u Z_{u-1}^u}\bigg|\mathscr{F}_t\right] - (1+K)\sum_{u=\tau}^{s} Z_t^u$$

where the last equality uses equation (6.2). Now using $Z_{u-1}^u = (1+r_u)^{-1} = B_{u-1}/B_u$ again, one obtains

$$\pi_t = \sum_{u=\tau}^{s} E_Q\left[\frac{B_t}{B_{u-1}}\bigg|\mathscr{F}_t\right] - (1+K)\sum_{u=\tau}^{s} Z_t^u$$
$$= \sum_{u=\tau}^{s} Z_t^{u-1} - (1+K)\sum_{u=\tau}^{s} Z_t^u$$
$$= Z_t^{\tau-1} - K\sum_{u=\tau}^{s-1} Z_t^u - (1+K)Z_t^s$$
$$= Z_t^{\tau-1} - \sum_{u=\tau}^{s} C_u Z_t^u$$

where the cash flow variables $C_u = K$ for $u = \tau, \ldots, s-1$ and $C_s = 1+K$ have been introduced. Thus the price of the forward swap is given by a simple present value calculation. In particular, in the case of an ordinary swap, for which $\tau = t+1$, the time-t price is

(6.44)
$$\pi_t = 1 - \sum_{u=t+1}^{s} C_u Z_t^u$$

This should be interpreted as one minus the time-t price of a coupon paying bond, a bond having face value one and coupon rate K.

The *forward swap rate* κ is that value of the fixed rate K which makes the time-t value of the forward swap zero, that is,

$$\kappa = \kappa(t,\tau,s) \equiv \frac{Z_t^{\tau-1} - Z_t^s}{Z_t^\tau + \ldots + Z_t^s}$$

The (ordinary) *swap rate* is simply the special case when $\tau = t+1$, that is, $\kappa(t, t+1, s)$

Example 6.1 (continued) With $K = 0.06$, $t = 0$, $\tau = 1$, and $s = 3$, the time-0 price of the payer swap is

$$\pi_0 = Z_0^0 - KZ_0^1 - KZ_0^2 - (1+K)Z_0^3$$
$$= 1 - 0.06(0.9434) - 0.06(0.893) - 1.06(0.844)$$
$$= -0.0049$$

in which case the price of the receiver swap is 0.0049. The swap rate is

$$\kappa = \frac{Z_0^0 - Z_0^3}{Z_0^1 + Z_0^2 + Z_0^3} = \frac{1 - 0.844}{0.9434 + 0.893 + 0.844} = 0.0582$$

A *payer swaption* is like a European call on the time $\tau - 1$ value of the corresponding payer forward start swap, where the exercise date is $\tau - 1$ and the exercise price is zero. A *receiver swaption* is defined in a similar manner with respect to the receiver swap. The payer and receiver swaptions thus have respective time-t prices ($t < \tau$)

$$E_Q\left[\frac{B_t}{B_{\tau-1}}\left(E_Q\left[\sum_{u=\tau}^{s}\frac{B_{\tau-1}}{B_u}(r_u - K)\bigg|\mathscr{F}_{\tau-1}\right]\right)^+\bigg|\mathscr{F}_t\right]$$

and

$$E_Q\left[\frac{B_t}{B_{\tau-1}}\left(E_Q\left[\sum_{u=\tau}^{s}\frac{B_{\tau-1}}{B_u}(K - r_u)\bigg|\mathscr{F}_{\tau-1}\right]\right)^+\bigg|\mathscr{F}_t\right]$$

Note the payer swaption price minus the receiver swaption price equals

$$E_Q\left[\frac{B_t}{B_{\tau-1}}E_Q\left[\sum_{u=\tau}^{s}\frac{B_{\tau-1}}{B_u}(r_u - K)\bigg|\mathscr{F}_{\tau-1}\right]\bigg|\mathscr{F}_t\right]$$
$$= E_Q\left[\frac{B_t}{B_{\tau-1}}\sum_{u=\tau}^{s}\frac{B_{\tau-1}}{B_u}(r_u - K)\bigg|\mathscr{F}_t\right] = E_Q\left[\sum_{u=\tau}^{s}\frac{B_t}{B_u}(r_u - K)\bigg|\mathscr{F}_t\right]$$

which is the time-t price of a forward start swap. Thus the following parity relationship holds:

(6.45) payer swaption − receiver swaption = forward swap

In view of equation (6.44), the time-($\tau - 1$) price of the payer swap is given by

$$\pi_{\tau-1} = 1 - \sum_{u=\tau}^{s} C_u Z_{\tau-1}^u$$

so another expression for the time-t price of the payer swaption is

$$E_Q\left[\frac{B_t}{B_{\tau-1}}\left(1 - \sum_{u=\tau}^{s} C_u Z_{\tau-1}^u\right)^+\bigg|\mathscr{F}_t\right]$$

This provides another interpretation:

(6.46) A payer swaption is the same as a put option on a coupon bond, where the exercise date is $\tau - 1$ and the exercise price is one. This coupon bond has face value one and coupon rate K.

Similarly, receiver swaptions can be interpreted as call options on coupon bonds.

Still another interpretation can be obtained:

(6.47) A payer (receiver) swaption is like a portfolio of call (respectively, put) options on the swap rate $\kappa(\tau-1,\tau,s)$.

To be precise, suppose for each time $u=\tau,\tau+1,\dots,s$, there is a call option with time-u payoff $[\kappa(\tau-1,\tau,s)-K]^+$. To see that this portfolio is like a payer swaption, consider its time-$(\tau-1)$ value, namely

$$E_Q\left[\sum_{u=\tau}^{s}\frac{B_{\tau-1}}{B_u}[\kappa(\tau-1,\tau,s)-K]^+\middle|\mathscr{F}_{\tau-1}\right]$$

It follows that the time-t value of this portfolio is

$$\begin{aligned}E_Q&\left[\frac{B_t}{B_{\tau-1}}\,E_Q\left[\sum_{u=\tau}^{s}\frac{B_{\tau-1}}{B_u}[\kappa(\tau-1,\tau,s)-K]^+\middle|\mathscr{F}_{\tau-1}\right]\middle|\mathscr{F}_t\right]\\&=E_Q\left[\frac{B_t}{B_{\tau-1}}[\kappa(\tau-1,\tau,s)-K]^+E_Q\left[\sum_{u=\tau}^{s}\frac{B_{\tau-1}}{B_u}\middle|\mathscr{F}_{\tau-1}\right]\middle|\mathscr{F}_t\right]\\&=E_Q\left[\frac{B_t}{B_{\tau-1}}[\kappa(\tau-1,\tau,s)-K]^+\sum_{u=\tau}^{s}Z^u_{\tau-1}\middle|\mathscr{F}_t\right]\end{aligned}$$

But $\kappa(\tau-1,\tau,s)[Z^\tau_{\tau-1}+\dots+Z^s_{\tau-1}]=Z^{\tau-1}_{\tau-1}-Z^s_{\tau-1}$ by the definition of the swap rate, so substituting this in the preceding expression yields

$$\begin{aligned}E_Q&\left[\frac{B_t}{B_{\tau-1}}\left(1-Z^s_{\tau-1}-K\sum_{u=\tau}^{s}Z^u_{\tau-1}\right)^+\middle|\mathscr{F}_t\right]\\&=E_Q\left[\frac{B_t}{B_{\tau-1}}\left(1-\sum_{u=\tau}^{s}C_uZ^u_{\tau-1}\right)^+\middle|\mathscr{F}_t\right]\end{aligned}$$

which is recognized to be the time-t price of the payer swaption by (6.46).

Example 6.4 With $K=0.06, t=1, \tau=2$, and $s=3$, the time-t price of the payer swap is

$$\begin{aligned}\pi_1&=1-KZ_1^2-(1+K)Z_1^3\\&=\begin{cases}1-0.06(0.9174)-1.06(0.8417)=\ \ 0.0528, & \omega=\omega_1,\omega_2\\1-0.06(0.9434)-1.06(0.8901)=-0.0001, & \omega=\omega_3,\omega_4\\1-0.06(0.9709)-1.06(0.9427)=-0.0576, & \omega=\omega_5,\omega_6\end{cases}\end{aligned}$$

The time-0 price of the payer swaption is

$$E_Q\left[\frac{B_0}{B_1}(\pi_1)^+\right]=(0.1667+0.1667)\left[\frac{1}{1.06}0.528\right]=0.0166$$

The time-0 price of the receiver swaption is

$$E_Q\left[\frac{B_0}{B_1}(-\pi_1)^+\right]=2(0.1667)\frac{0.0001}{1.06}+2(0.1667)\frac{0.0576}{1.06}=0.0181$$

Note that the time-0 price of the forward start swap is

$$\frac{0.1667+0.1667}{1.06}[0.0528-0.0001-0.0576]=-0.0015$$

so, indeed, the parity relationship (6.45) is satisfied.

To verify interpretation (6.47), note that we have

$$\begin{aligned}\kappa=\kappa(\tau-1,\tau,s)=\kappa(1,2,3)&=\frac{Z_1^1-Z_1^3}{Z_1^2+Z_1^3}\\&=\begin{cases}\frac{1-0.8417}{0.9174+0.8417}=0.09, & \omega=\omega_1,\omega_2\\\frac{1-0.8901}{0.9434+0.8901}=0.06, & \omega=\omega_3,\omega_4\\\frac{1-0.9427}{0.9709+0.9427}=0.03, & \omega=\omega_5,\omega_6\end{cases}\end{aligned}$$

so

$$(\kappa-K)^+=(\kappa-0.06)^+=\begin{cases}0.03, & \omega=\omega_1,\omega_2\\0, & \text{otherwise}\end{cases}$$

The portfolio of interest has two options, one paying off at time $t=2$, the other paying off at $t=3$. Each pays 0.03 dollars in the event $A\equiv\{\omega_1,\omega_2\}$ but pays zero otherwise. The time-0 present value of this cash flow is

$$\begin{aligned}E_Q&\left[\frac{0.03}{B_2}1_A\right]+E_Q\left[\frac{0.03}{B_3}1_A\right]\\&=\frac{2(0.1667)(0.03)}{1.06(1.09)}+\frac{0.1667(0.03)}{1.06(1.09)(1.1)}+\frac{0.1667(0.03)}{1.06(1.09)(1.08)}=0.0166\end{aligned}$$

which equals the time-0 price of the payer swaption, in accordance with (6.47).

Exercise 6.23 For the model in example 6.2 with $K=0.06$, $\tau=1$, and $s=3$, compute the time-0 prices of the payer and receiver swaps, and compute the swap rate at $t=0$.

Exercise 6.24 For the model in example 6.1 with $K=0.06$, $\tau=2$, and $s=3$

(a) Compute the time-1 prices of the payer swap.
(b) Compute the time-0 price of the payer forward start swap.
(c) Compute the time-0 price of the payer swaption.

(d) Compute the time-0 price of the receiver swaption, and verify parity relationship (6.45).
(e) Compute the swap rate $\kappa(1,2,3)$.
(f) Verify (6.47) by specifying the appropriate option portfolio and computing its time-0 value.

Exercise 6.25 Show that the time-t price of a payer forward start swap *settled in advance* is

$$\pi_t = E_Q\left[\sum_{u=\tau}^{s} \frac{B_t}{B_{u-1}}(r_u - K)\middle|\mathscr{F}_t\right], \quad t < \tau \leqslant s \leqslant T$$

6.7 Caps and Floors

A *caplet* is a European call option on the spot interest rate r at a fixed point in time. As with swaps, caplets can be settled in arrears or in advance. In the former case, the time-τ payoff of a caplet is simply $(r_\tau - K)^+$, where K is the specified strike or exercise price. For $t \leqslant \tau$ the time-t price of this caplet is

$$\pi_t = B_t E_Q[(r_\tau - K)^+ / B_\tau | \mathscr{F}_t] = Z_t^\tau E_\tau[(r_\tau - K)^+ | \mathscr{F}_t]$$

A *cap* is a strip of caplets, all having a common exercise price and with one for each time period in an interval of time. In general, some caplets in the cap will pay off and others will not, depending on whether the spot rate exceeds the strike. As with swaps, there are both ordinary and forward start caps, depending on whether the initial caplet corresponds to the current spot rate. The time-t price of a forward start cap settled in arrears is

(6.48) $$\pi_t = B_t \sum_{u=\tau}^{s} E_Q[(r_u - K)^+ / B_u | \mathscr{F}_t], \quad t < \tau \leqslant s \leqslant T$$

The time-t price of an ordinary cap is given by the same formula with $\tau = t + 1$.

Example 6.1 (continued) Suppose the exercise price $K = 0.06$ and consider an ordinary cap settled in arrears. The payoff of the $\tau = 1$ caplet is identical to zero, so its time-0 price is zero. The payoff of the $\tau = 2$ caplet is

$$(r_2(\omega) - 0.06)^+ = \begin{cases} 0.03, & \omega = \omega_1, \omega_2 \\ 0, & \text{otherwise} \end{cases}$$

so its time-0 price is

$$E_Q[(r_2 - 0.06)^+ / B_2] = (0.1839 + 0.1161)\frac{0.03}{(1.06)(1.09)} = 0.0078$$

The payoff of the $\tau = 3$ caplet is

$$(r_3(\omega) - 0.06)^+ = \begin{cases} 0.04, & \omega = \omega_1 \\ 0.02, & \omega = \omega_2 \\ 0.01, & \omega = \omega_3 \\ 0, & \text{otherwise} \end{cases}$$

so its time-0 price is

$$E_Q[(r_3 - 0.06)^+ / B_3] = \frac{0.1839(0.04)}{1.06(1.09)(1.1)} + \frac{0.1161(0.02)}{1.06(1.09)(1.08)} + \frac{0.1517(0.01)}{1.06(1.06)(1.07)} = 0.0091$$

Thus the time-0 price of the corresponding cap with $\tau = 1$ and $s = 3$ is $0.0078 + 0.0091 = 0.0169$.

Floorlets are defined in the same manner as caplets, only they are put options on the spot rate. Similarly, *floors* are strips of floorlets and thus are analogous to caps. In particular, the time-t price of a forward start floor settled in arrears is

(6.49) $$\pi_t = B_t \sum_{u=\tau}^{s} E_Q[(K - r_u)^+ / B_u | \mathscr{F}_t], \quad t < \tau \leqslant s \leqslant T$$

Comparing this with (6.43) and (6.48) immediately gives the following parity relationship:

(6.50) The price of a cap minus the price of a floor equals the price of a swap.

Example 6.1 (continued) Consider the floor that is settled in arrears with $K = 0.06, \tau = 1$, and $s = 3$. Since $r_1 = 0.06$, the $\tau = 1$ floorlet has time-0 price equal to zero. The $\tau = 2$ floorlet has time-0 price

$$E_Q[(0.06 - r_2)^+ / B_2] = (0.2582 + 0.1418)\frac{0.03}{(1.06)(1.03)} = 0.011$$

The $\tau = 3$ floorlet has time-0 price

$$\frac{0.1483(.01)}{1.06(1.06)(1.05)} + \frac{0.2582(.02)}{1.06(1.03)(1.04)} + \frac{0.1418(.04)}{1.06(1.03)(1.02)} = 0.0108$$

Thus the time-0 price of the floor is $0.011 + 0.0108 = 0.0218$. Note this is consistent with parity relationship (6.50), because the price of the corresponding cap is 0.0169, whereas (see section 6.6) -0.0049 is the price of the corresponding swap.

A *caption* is a put or call option whose underlying security is a forward cap. A *floortion* is a put or call option whose underlying security

is a forward floor. These are examples of what are called *compound options*, that is, options whose underlying securities are themselves options.

Normally the exercise date of the caption or floortion is prior to the time associated with the initial caplet or floorlet in the underlying. Moreover, the strike of the caption or floortion is generally not the same as the strike of the cap or floor.

The computation of the price of a caption or floortion is best viewed as a two-step procedure. First you compute the probability distribution of the caption or floortion payoff under either the risk neutral or the forward risk adjusted probability measure. In other words, you first compute the probability distribution of the contingent claim. Then you compute in a standard way the price of this contingent claim. This is illustrated in the following example.

Example 6.1 (continued) First consider a forward start cap with $K = 0.06, \tau = 2$ and $s = 3$. The $\tau = 2$ caplet has time-1 price

$$B_1 E_Q[(r_2 - 0.06)^+/B_2|\mathscr{F}_1] = \begin{cases} 1(0.03)/1.09 = 0.0275, & \omega = \omega_1, \omega_2 \\ 0, & \text{otherwise} \end{cases}$$

The $\tau = 3$ caplet has time-1 price

$$B_1 E_Q[(r_3 - 0.06)^+/B_3|\mathscr{F}_1]$$

$$= \begin{cases} \frac{0.613(0.04)}{1.09(1.1)} + \frac{0.387(0.02)}{1.09(1.08)} = 0.0271, & \omega = \omega_1, \omega_2 \\ \frac{0.5057(0.01)}{1.06(1.07)} = 0.0045, & \omega = \omega_3, \omega_4 \\ 0, & \omega = \omega_5, \omega_6 \end{cases}$$

where 0.613 equals the conditional probability $Q(\omega_1|\{\omega_1, \omega_2\})$, and so forth. Thus the time-1 price of this forward start cap is

$$\pi_1 = \begin{cases} 0.0275 + 0.0271 = 0.0546, & \omega = \omega_1, \omega_2 \\ 0.0045, & \omega = \omega_3, \omega_4 \\ 0, & \omega = \omega_5, \omega_6 \end{cases}$$

Notice that the time-0 price of this cap is

$$\pi_0 = E_Q[\pi_1/B_1] = \frac{0.3(0.0546)}{1.06} + \frac{0.3(0.0045)}{1.06} = 0.0168$$

as was computed earlier in a different manner.

Now consider a caption, in particular, a put on this cap with strike 0.02 and exercise date $\tau = 1$. Its time-1 payoff is

$$(0.02 - \pi_1)^+ = \begin{cases} 0, & \omega = \omega_1, \omega_2 \\ 0.0155, & \omega = \omega_3, \omega_4 \\ 0.02, & \omega = \omega_5, \omega_6 \end{cases}$$

in which case the time-0 price of the caption is

$$E_Q[(0.02 - \pi_1)^+/B_1] = \frac{0.3(0.0155)}{1.06} + \frac{0.4(0.02)}{1.06} = 0.0119$$

Similarly, the time-0 price of the caption which is a call having the same strike and exercise date is

$$E_Q[(\pi_1 - 0.02)^+/B_1] = \frac{0.3(0.0346)}{1.06} = 0.0098$$

Exercise 6.26 Consider the model in example 6.2.

(a) Compute the time-0 and time-1 prices of the forward start cap that is settled in arrears with $K = 0.06, \tau = 2$, and $s = 3$.
(b) Compute the time-0 and time-1 prices of the forward start floor that is settled in arrears with the same parameters.
(c) Verify that the time-0 prices in (a) and (b) satisfy the parity relationship (6.50).
(d) Compute the time-0 prices of the put and call captions, where the underlying is as in (a), the strike is 0.01, and the exercise date is $\tau = 1$.
(e) Compute the time-0 prices of the put and call floortions, where the underlying is as in (b), the strike is 0.01, and the exercise date is $\tau = 1$.

7 Models With Infinite Sample Spaces

7.1 Finite Horizon Models

The fundamental theorem of asset pricing says that there are no arbitrage opportunities if and only if there exists a risk neutral probability measure. In earlier chapters this principle was shown true for single and multiperiod models under the critical assumption that the underlying sample space Ω has a finite number of elements. This assumption is crucial, because it enables one to apply simple results of linear programming or, more generally, to use simple versions of the separating hyperplane theorem for problems posed in terms of finite dimensional spaces. But when the sample space Ω has a countably infinite or an uncountably infinite number of elements, the space of random variables representing terminal wealth will be a space of infinite dimension. There are certainly separating hyperplane theorems for these infinite dimensional settings, but a straightforward application of such a theorem will break down owing to technical complications. A more delicate analysis is required.

It turns out the fundamental theorem of asset pricing remains true in the case of infinite sample spaces Ω provided the number T of trading periods is finite. The purpose of this section is to establish this result. As will be seen in the following section, however, the theorem breaks down when the number T of trading periods is infinite.

Some results in this chapter are more technical than much of the earlier chapters. To begin with, in progressing to an infinite sample space it is necessary to generalize the concept of the filtration that is used as the information submodel. Recall that a collection $\mathscr{F}$ of subsets of Ω is called an *algebra* on Ω if

1 $\Omega \in \mathscr{F}$
2 $F \in \mathscr{F} \Rightarrow F^c = \Omega \backslash F \in \mathscr{F}$
3 F and $G \in \mathscr{F} \Rightarrow F \cup G \in \mathscr{F}$.

The collection $\mathscr{F}$ is called a *σ-algebra* on Ω if, in addition

4 $F_1, F_2, \ldots \in \mathscr{F} \Rightarrow \bigcup_{n=1}^{\infty} F_n \in \mathscr{F}$.

In words, (4) says that countable unions of subsets in $\mathscr{F}$ are also in $\mathscr{F}$. If Ω is finite, then (4) is superfluous, because Ω has only finitely many subsets. But if Ω has infinitely many subsets then (3) does not imply (4). Condition (4) is necessary for what needs to be done.

Our aim is to use σ-albegras as models for the information that is known to the investors at the various points in time. Recall that algebras were used for this purpose in the finite sample space context. Moreover, recall that when Ω is finite, algebras on Ω are equivalent to partitions of Ω, and this led to a very intuitive interpretation of algebras as models of the information known to the investors. Unfortunately, the analogous result does not hold for σ-algebras on Ω when Ω has an infinite number of elements. Recall from chapter 3 that for $A \subset \Omega$ to be an element of the partition $\mathscr{P}$ corresponding to $\mathscr{F}$, one must have $A \in \mathscr{F}$ as well as $\{B \in \mathscr{F},\ A \neq B, \text{ and } B \subset A\} \Rightarrow B = \emptyset$. But applying this requirement to infinite sample spaces produces silly, worthless results. For instance, if $\Omega = [0, 1]$ and $\mathscr{F}$ includes all the open intervals of the form (a, b), where a and b are rational numbers, then this requirement implies the corresponding partition includes all the rational numbers of the unit interval as distinct elements.

Nevertheless, we will use σ-algebras to model the information known to the economic agents at individual points in time. The economic intuition to justify this approach is dubious, since the partition interpretation was abandoned. We must soldier on, being content with the view that σ-algebras are natural extensions of the concept that made good sense when Ω was finite.

As with the case of a finite sample space, the information flow in the securities market will be modeled by a *filtration* $\mathbb{F} = \{\mathscr{F}_t;\ t = 0, 1, \ldots T\}$, where $\{\mathscr{F}_t\}$ is an increasing sequence of σ-albegras. In particular, $\mathscr{F}_{t-1} \subseteq \mathscr{F}_t$ for $t = 1, \ldots, T$, because the investors learn as time goes on. The probability measure P is such that the probability $P(A)$ is well defined for each $A \in \mathscr{F}_T$.

The random variable X is said to be *measurable* with respect to the σ-algebra $\mathscr{F}$ if, for every real number x, the subset $\{\omega \in \Omega : X(\omega) \leqslant x\}$ is an element of the σ-algebra $\mathscr{F}$. In this case one writes $X \in \mathscr{F}$. The stochastic process $X = \{X_t; t = 0, \ldots, T\}$ is said to be *adapted* if $X_t \in \mathscr{F}_t$ for all t. The stochastic process $H = \{H_t; t = 1, \ldots, T\}$ is said to be *predictable* if $H_t \in \mathscr{F}_{t-1}$ for all t.

At this point it is convenient to introduce a new and rather technical concept. When speaking about equations, inequalities, and the like involving random variables, it is often appropriate to include the phrase *almost surely*. This means that the equation or whatever holds for all $\omega \in \Omega$ except, possibly, for some $\omega \in A$, where A is some event in $\mathscr{F}_T$ such that $P(A) = 0$. In other words, although the relationship in question might not hold for all $\omega \in \Omega$, the only exceptions are inconsequential and can be ignored. *Almost surely* is often abbreviated *a.s.*

The rest of the securities market model is largely the same as before. There is a *bank account* process $B = \{B_t; t = 0, 1, \ldots, T\}$, an adapted, non-decreasing stochastic process with $B_o = 1$. There are N *risky security*

processes $S_n = \{S_n(t); t = 0, 1, \ldots, T\}$, where S_n is a non-negative, adapted stochastic process for each $n = 1, 2, \ldots, N$. And there are N *discounted price processes* $S_n^* = \{S_n^*(t);\ t = 0, 1, \ldots, T\}$, where $S_n^*(t) \equiv S_n(t)/B_t$ for all t.

We are now ready to tackle the fundamental theorem of asset pricing. This will be done after first presenting a preliminary result which is so technical that its explanation will be omitted.[1] Here Y and Z are each random variables taking values in $\mathbb{R}^N$, $\| Y \|$ denotes the Euclidean norm of the vector Y, and $Z \cdot Y$ denotes the inner product of the vectors Y and Z.

(7.1) Suppose $\mathscr{G}$ and $\mathscr{H}$ are two σ-algebras with $\mathscr{G} \subseteq \mathscr{H} \subseteq \mathscr{F}_T$. Let $Y \in \mathscr{H}$ be such that

(7.2) $$\{Z \in \mathscr{G} \text{ and } Z \cdot Y \geqslant 0 \text{ a.s.}\} \Rightarrow \{Z \cdot Y = 0 \text{ a.s.}\}$$

Then there exists a scalar-valued random variable $D \in \mathscr{H}$ such that:

(7.3) $$0 < D \leqslant 1, \text{a.s.}$$

(7.4) $$E\,[D \,\|\, Y \,\|] < \infty, \text{ and}$$

(7.5) $$E[YD|\mathscr{G}] = 0$$

In words, (7.2) says that the only random variables $Z \in \mathscr{G}$ satisfying the inequality actually satisfy the equality. Shortly it will be seen that this is the same as saying there are no one-period arbitrage opportunities. The random variable D will play the role of a state price density and will be used in the construction of the risk neutral probability measure.

The trading strategy H, the value process V, the discounted value process V^*, and the discounted gains process G^* are defined in exactly the same manner as in the case of a finite sample space Ω (see section 3.1). Moreover, arbitrage opportunities are also defined the same way: the self-financing trading strategy H is an *arbitrage opportunity* if (1) $V_0 = 0$, (2) $V_T \geqslant 0$, and (3) $E[V_T] > 0$. Just as with (3.17), an equivalent condition for H to be an arbitrage opportunity is for the discounted gains process to satisfy $G_T^* \geqslant 0$ and $E[G_T^*] > 0$.

In view of this latter characterization of arbitrage opportunities, it is not surprising that the absence of arbitrage opportunities implies a condition that resembles (7.2). To prepare for this, let $S^* = \{S_t^*; t = 0, \ldots, T\}$ denote the $\mathbb{R}^N$-valued process whose nth component is S_n^*, the discounted price process for the nth risky security, $n = 1, \ldots, N$. There should not be any confusion about whether the subscripts here represent time t or security n. It is convenient to denote $\Delta S_t^* = S_t^* - S_{t-1}^*$, a $\mathbb{R}^N$-valued random variable.

(7.6) If there are no arbitrage opportunities, then for all $t \in \{1, 2, \ldots, T\}$ and all $\mathbb{R}^N$-valued random variables Z,

(7.7) $$\{Z \in \mathscr{F}_{t-1} \text{ and } Z \cdot \Delta S_t^* \geqslant 0 \text{ a.s.}\} \Rightarrow \{Z \cdot \Delta S_t^* = 0 \text{ a.s.}\}$$

Principle (7.6) can be verified by contradiction, because if there exists some t and some $Z = (Z_1, \ldots, Z_N) \in \mathscr{F}_{t-1}$ such that $Z \cdot \Delta S_t^* \geqslant 0$ *a.s.* and $P(Z \cdot \Delta S_t^* > 0) > 0$, then one can construct an arbitrage opportunity, as will now be shown.

Let $A \in \mathscr{F}_{t-1}$ denote the set $\{\omega \in \Omega : P(Z \cdot \Delta S_t^* > 0 | \mathscr{F}_{t-1})(\omega) > 0\}$, and note that $P(A) > 0$ by assumption. The arbitrage opportunity H is obtained by taking $H_n(s)(\omega) = 0$ for all $s < t$, all $\omega \in \Omega$, and $n = 0, 1, \ldots, N$; by taking

$$H_n(t)(\omega) = \begin{cases} Z_n(\omega), & \omega \in A, n = 1, \ldots, N \\ -Z \cdot S_{t-1}^*(\omega), & \omega \in A, n = 0 \\ 0, & \omega \in A^C \end{cases}$$

and by taking

$$H_n(s)(\omega) = \begin{cases} V_t(\omega), & \text{n=0 and } \omega \in A \\ 0, & \text{otherwise} \end{cases}$$

for $s = t+1, \ldots, T$. Clearly H is predictable and $V_0 = 0$. It is left for the reader to verify that H is self-financing (exercise 7.1). This strategy takes the time-t wealth V_t and holds it in the bank account, so $V_T \geqslant 0$ *if* $V_t \geqslant 0$. But $V_t(\omega) = Z \cdot \Delta S_t^*(\omega) \geqslant 0$ if $\omega \in A$, whereas $V_t(\omega) = 0$, otherwise, so indeed $V_T \geqslant 0$. Moreover, $E[V_T] > 0$ if $E[V_t] > 0$. But

$$E[V_t] = E[1_A Z \cdot \Delta S_t^*] = E[1_A E[Z \cdot \Delta S_t^* | \mathscr{F}_{t-1}]] > 0$$

so all this implies H is an arbitrage opportunity, which is a contradiction.

Principles (7.1) and (7.6) can now be used to show that the absence of arbitrage opportunities implies the existence of a risk neutral probability measure. A *risk neutral probability measure* is defined in the same way as in the case of a finite sample space; it is a probability measure Q, equivalent to P, under which the discounted price of each risky security is a martingale. By *equivalent*, one means that for each event $A \in \mathscr{F}_T$ the probability $P(A) = 0$ if and only if $Q(A) = 0$. In words, there is agreement between the two probability measures about which events can and cannot happen.

To see why no arbitrage implies the existence of a risk neutral probability measure Q, I will indicate how to construct this measure. Begin by setting $\mathscr{F}_{T+1} = \mathscr{F}_T, S_{T+1}^* = S_T^*, D_{T+1} = 1$, and $Y_{T+1} = 0$. The scalar-valued random variables $D_1, \ldots, D_T$ and the $\mathbb{R}^N$-valued random variables $Y_1, \ldots, Y_T$ are now defined recursively, backwards in time. For example, suppose $D_{t+1}, \ldots, D_T$ and $Y_{t+1}, \ldots, Y_T$ have been defined such that

(7.8) $$D_k \text{ is } \mathscr{F}_k \text{ measurable, } t+1 \leqslant k \leqslant T+1$$

(7.9) $$Y_k = \Delta S_k^* E[D_{k+1} \ldots D_{T+1} | \mathscr{F}_k] \; a.s., \; t+1 \leqslant k \leqslant T$$

(7.10) $$0 < D_k \leqslant 1 \; a.s., \; t+1 \leqslant k \leqslant T+1$$

(7.11) $$E\,[D_K \| Y_k \|] < \infty, \; t+1 \leqslant k \leqslant T, \text{and}$$

(7.12) $$E\,[Y_k D_k | \mathscr{F}_{k-1}] = 0 \; a.s., \; t+1 \leqslant k \leqslant T$$

Now use principle (7.1) with $\mathscr{G} = \mathscr{F}_{t-1}$, $\mathscr{H} = \mathscr{F}_t$, and $Y = Y_t$, where Y_t is as in (7.9) with $k = t$. Note that $Y_t \in \mathscr{F}_t$. Moreover, because of (7.6), property (7.2) holds. Thus by (7.1) there exists a scalar-valued random variable D_t such that equations (7.8), (7.10), (7.11), and (7.12) all hold with $k = t$.

Having specified $D_1, \ldots, D_T$, the next step is to define a few more items. Set

$$D_0 \equiv \frac{1}{1 + \|S_0^*\|} \quad \text{and} \quad D \equiv D_0 D_1 \ldots D_T$$

Note that $0 < D \leqslant 1$ a.s. In addition, define what will turn out to be the risk neutral probability measure by taking

$$Q(A) \equiv \frac{E[D1_A]}{E[D]}, \quad \text{all events } A \in \mathscr{F}_T$$

that is,

$$E_Q[X] \equiv \frac{E[DX]}{E[D]}, \quad \text{all random variables } X \in \mathscr{F}_T$$

Clearly Q is a probability measure that is equivalent to P. It remains to show that the discounted risky security prices are martingales under Q. There are two parts to this. The first is to show the usual conditional expected value relationship. The other is to show that the discounted risky price process is *integrable* under Q, that is, to show that various expected values of this process are well-defined and finite.

To verify Q-integrability, we have

$$\begin{aligned} E_Q\left[\|S_0^*\|\right] &= \frac{1}{E[D]} E\left[D\|S_0^*\|\right] = \frac{1}{E[D]} E\left[D_0 D_1 \ldots D_T \|S_0^*\|\right] \\ &\leqslant \frac{1}{E[D]} E\left[D_0 \|S_0^*\|\right] \leqslant \frac{1}{E[D]} < \infty \end{aligned}$$

We also have

$$\begin{aligned} E_Q\left[\|S_t^* - S_{t-1}^*\|\right] &= \frac{1}{E[D]} E\left[D\|S_t^* - S_{t-1}^*\|\right] \\ &= \frac{1}{E[D]} E\left[D_0 \ldots D_t \|S_t^* - S_{t-1}^*\| D_{t+1} \ldots D_T\right] \\ &= \frac{1}{E[D]} E\left[D_0 \ldots D_t \|S_t^* - S_{t-1}^*\| E\left[D_{t+1} \ldots D_T | \mathscr{F}_t\right]\right] \\ &= \frac{1}{E[D]} E\left[D_0 \ldots D_t \|Y_t\|\right] \\ &\leqslant \frac{1}{E[D]} E\left[D_t \|Y_t\|\right] \end{aligned}$$

But this last expression is finite by (7.11), so S^* is Q-integrable.

It remains to show $E_Q[S_t^* | \mathscr{F}_{t-1}] = S_{t-1}^*$ for all t, that is, to show $E_Q[\Delta S_t^* | \mathscr{F}_{t-1}] = 0$ for all t. We have

$$\begin{aligned} E_Q[\Delta S_t^* | \mathscr{F}_{t-1}] &= \frac{1}{E[D]} E[D \Delta S_t^* | \mathscr{F}_{t-1}] \\ &= \frac{1}{E[D]} E[D_0 \ldots D_{t-1} D_t (\Delta S_t^*) D_{t+1} \ldots D_T | \mathscr{F}_{t-1}] \\ &= \frac{D_0 \ldots D_{t-1}}{E[D]} E[D_t (\Delta S_t^*) E[D_{t+1} \ldots D_T | \mathscr{F}_T] | \mathscr{F}_{t-1}] \\ &= \frac{D_0 \ldots D_{t-1}}{E[D]} E[D_t Y_t | \mathscr{F}_{t-1}] \end{aligned}$$

But this last expression equals 0 by (7.12), so S^* is indeed a martingale under Q.

In summary, if there are no arbitrage opportunities, then there exists a risk neutral probability measure. The converse is immediate, by the same argument as used for the finite sample space case. Hence we have established that the following is true even if the sample space Ω has infinitely many elements:

(7.13) *Fundamental Theorem of Asset Pricing* Suppose the number of trading periods T is finite. Then there are no arbitrage opportunities if and only if there exists a risk neutral probability measure.

Exercise 7.1 The statement was made that the arbitrage opportunity constructed in the explanation of (7.6) is a self-financing trading strategy. Verify this assertion.

Exercise 7.2 Suppose there is a single security S with $\Delta S_t = \exp\{\sigma W_t + \mu\}$, where $\{W_t\}$ is a sequence of independent standard normal random variables and where μ and σ are positive constants. In addition, suppose the spot interest rate is the constant $r \geqslant 0$. Derive a risk neutral probability measure, first for the case $T = 1$, then for the general case $T < \infty$.

7.2 Infinite Horizon Models

Finite horizon models with infinite sample spaces are not much different from finite horizon models with finite sample spaces, because the fundamental theorem of asset pricing holds in both cases. But this theorem is not quite true when there is an infinite number of trading periods, and there are also significant modeling issues associated with admissibility of trading strategies. This section will examine these and related matters.

First of all, what is meant by an infinite horizon model? For finite horizon models it was tacitly assumed that the time index t keeps track of both the number of periods as well as the elapsed time by some unit of measure such

as months or years. This implies the time periods are of equal durations, such as one year. Retaining this set-up for an infinite horizon problem leads to the choice $T = \infty$, meaning there is a (countably) infinite number of trading periods, all with the same length.

There is an alternative approach, however, that is useful for some purposes: there is a finite planning horizon, but there is also a (countably) infinite number of trading periods before the planning horizon. Here, of course, the various periods have different lengths, as measured by clock time. For instance, with a planning horizon of one year, period t would last for $(1/2)^t$ years, $t = 1, 2, \ldots$ Either way (and, admittedly, the terminology is not entirely consistent), the defining feature of *infinite horizon models* is the (countably) infinite number of trading periods; whether the planning horizon measured in clock time is finite and whether all the periods are of the same length of clock time are unimportant issues for the purposes of this section. Throughout, one should think of t as the counter for the number of periods, thereby suggesting $T = \infty$.

Building the security price processes for the infinite horizon model presents no great difficulties. Needless to say, the sample space Ω will necessarily be infinite. With the specification of the probability space $(\Omega, \mathscr{F}, P)$, the filtration model of the information, $\mathbb{F} = \{\mathscr{F}_t; t = 0, 1, \ldots\}$, will be a collection of σ-algebras with $\mathscr{F}_t \subseteq \mathscr{F}_{t+1} \subseteq \ldots \subseteq \mathscr{F}$ for all t. The security price processes will then be, as usual, non-negative, adapted stochastic processes.

When it comes to the specification of the trading strategies, however, we quickly encounter a complication. As usual, we will want to require the trading strategies to be predictable stochastic processes, for this will enable the investors to use all the past and present available information without being able to look into the future. But without additional restrictions, the opportunity to trade infinitely many times will allow investors to make arbitrage profits, even with perfectly reasonable stock price models. This will be illustrated in the following example, which deals with the very simple binomial stock price model.

Example 7.1 Consider a simple binomial stock price model where the "up" factor $u = 1.1$,, the "down" factor $d = 0.9$, and the riskless interest rate $r = 0$. The objective here is to describe an arbitrage opportunity where you start with zero dollars and you are certain to end up with \$1. The probability of an up-move is strictly between 0 and 1; the exact value is unimportant. But note there is a risk neutral probability measure for this model; it is the one consistent with equal conditional probabilities for up and down moves over one period.

The idea will be to use a 'doubling' strategy where you start out at time $t = 0$ borrowing \$10, say, investing all of this in the stock. If the stock goes up in the first period, then your investment becomes \$11, so you repay your \$10 loan and you take the remaining \$1 and run. On the other hand, if the stock goes down, then your stock investment becomes worth \$9; since you owe \$10, this puts you \$1 in debt. You can "recover" by borrowing enough money to double the initial \$10 investment in the stock. In particular, you borrow an additional \$11, raising the total loan to \$21 and making the total investment in the stock equal to \$20. If the stock goes up over the second period, then your stock investment becomes \$22, so you repay your \$21 loan and you take the remaining \$1 and run. But if the stock goes down, then your stock investment becomes \$18, leaving you \$3 in debt, requiring you to "double-up" at least one more time.

In general, if there is at least one up move during the first t periods, then you will have realized your desired \$1 profit and you will have terminated all borrowing and trading by time t. On the other hand, if the stock goes down each of the first t periods, then after t periods you will be $1 - 2^t$ in debt, you will owe $11(2)^{t-1} - 1$ on your loan, and your current investment in the stock will be worth $9(2)^{t-1}$ (check: $1 - 2^t = 9(2)^{t-1} - [11(2)^{t-1} - 1]$). In this latter case then at time t you will borrow an additional $11(2)^{t-1}$ dollars to increase your stock investment to $10(2)^t$ dollars and keep on going.

Now if there is only a finite number T of periods, then with a positive probability the stock will have gone down every period and you will have ended up $1 - 2^T$ dollars in debt, in which case this would not be an arbitrage opportunity. But with an infinite number of trading periods, the probability of always going down and thus ending up in debt is equal to zero. In other words, under the doubling strategy the probability of ending up with \$1 as your final wealth is equal to one. This is an arbitrage opportunity.

So what is the problem? Is our notion of an arbitrage opportunity improper for the case of an infinite horizon model? It is questionable to talk about being sure of having a positive wealth infinitely far out in the future when you run the risk of being in debt at every finite time, but keep in mind that the infinite horizon model can be associated with a planning horizon of finite clock time, with an infinite number of trading periods of varying length. With this latter perspective we would be referring to a positive wealth at a finite distance in the future. Our notion of arbitrage opportunities is not the problem.

The problem with example 7.1 is that the specified trading strategy is unrealistic. In the first place, there is no lower bound on the amount the investor could be in debt (this could be $1 - 2^t$ after t periods), nor is there an upper bound on the amount (which could be $11(2)^{t-1} - 1$ dollars after t periods) of the loan. Furthermore, there is no upper bound on the shares of stock that the investor might need to own (the dollars invested grow without bound as the price per share drops, so the number of shares owned will grow without bound as well). These situations are unrealistic from the modeling and economic points of view.

It is reasonable to make an assumption that would rule out these unrealistic situations. For example, one could stipulate that the trading strategies

(i.e., the number of shares long or short) are bounded. Or one could stipulate that there is a lower bound on the wealth of the investor. Assumptions like these will rule out the doubling strategies that produce arbitrage opportunities. This is the approach I will take.

I now return to the fundamental theorem of asset pricing, which says there exists a risk neutral probability measure if and only if there are no arbitrage opportunities. In order to discuss this, and in view of the preceding discussion, it is necessary to give a careful definition about arbitrage opportunities in the case of an infinite horizon model (the definition of the risk neutral probability measure is no different than in the finite horizon case). A predictable trading strategy H will be said to be *admissible* if there exists a scalar $m < \infty$ such that the time-t wealth $V_t \geqslant -m$ for all t (this inequality holds with probability one; note admissibility rules out doubling strategies). The admissible, self-financing trading strategy H is said to be an *arbitrage opportunity* if (recall $V_t^* = V_t/B_t$ is the time-t discounted value of the portfolio):

1 $V_0 = V_0^* = 0$
2 There exists a random variable $V^* \in \mathcal{F}$ such that $V_t^* \to V^*$ as $t \to \infty$ (that is, $P(V_t^* \to V^*) = 1$)
3 $V^* \geqslant 0$
4 $E[V^*] > 0$

Thus (1), (3), and (4) are the same as in the finite horizon case, except that here they involve the discounted wealth V^* after an infinite number of trading periods, with V^* given in (2).

It is now easy to explain the following:

(7.14) If there exists a risk neutral probability measure Q, then there are no arbitrage opportunities.

To see this, suppose H is an admissible trading strategy with V^* as in (2) and with $V_0^* = 0$. Then just as in the finite horizon case, $t \to V_t^*$ is a martingale under Q, so $E_Q[V_t^*] = V_0^* = 0$ for all t. It follows from Fatou's Lemma (a technical convergence theorem for sequences of random variables) that $E_Q[V^*] \leqslant 0$, so V^* cannot satisfy both (3) and (4). Thus H cannot be an arbitrage opportunity.

Unfortunately, the converse of (7.14), which together with (7.14) would comprise the Fundamental Theorem of Asset Pricing, is not true for general infinite horizon models. This will be illustrated with the following example.

Example 7.2 Consider a securities market model with one risky security S, $r = 0$, and a countable sample space $\Omega = \{1, 2\ldots\}$. Set $S_0 = 1$ and, for all $t \geqslant 1$ and all $\omega \in \Omega$, set

$$S_t(\omega) = \begin{cases} (1/2)^t & t < \omega \\ (\omega^2 + 2\omega + 2)(1/2)^\omega, & t \geqslant \omega \end{cases}$$

This means that in state ω the price falls by 50 per cent per period for $\omega - 1$ consecutive periods, from time $\omega - 1$ to time ω the price increases by $(\omega^2 + 2\omega)(1/2)^\omega$, and from then on the price is constant. In other words,

$$\Delta S_t(\omega) = S_t(\omega) - S_{t-1}(\omega) = \begin{cases} -(1/2)^t, & t < \omega \\ (\omega^2 + 2\omega)(1/2)^\omega, & t = \omega \\ 0, & t > \omega \end{cases}$$

Let H_t denote the number of shares of the risky security held from time $t - 1$ to time t. Given the nature of the price process, nothing is lost by restricting attention to trading strategies where $\{H_t\}$ is a sequence of real numbers. In order to avoid doubling strategies, it will also be assumed that the sequences $\{H_t\}$ are bounded. Of course, the trading strategies must be self-financing. Hence the admissible trading strategies are fully described by the initial wealth V_0 and the bounded sequence $\{H_t\}$ of real numbers.

If $t < \omega$, then during period t the portfolio loses $(1/2)^t H_t$ in value, whereas if $t = \omega$, then the portfolio gains $(\omega^2 + 2\omega)(1/2)^\omega H_t$. Of course, if $t > \omega$, then the portfolio remains constant in value, even though the value of H_t might be non-zero. Hence the time t value of the portfolio under admissible trading strategy $(V_0, \{H_t\})$ is

$$V_t(\omega) = \begin{cases} V_0 - \sum_{s=1}^{t} (1/2)^s H_s & t < \omega \\ V_0 - \sum_{s=1}^{\omega-1} (1/2)^s H_s + (\omega^2 + 2\omega)(1/2)^\omega H_\omega, & t \geqslant \omega \end{cases}$$

It is easy to verify that the value process V is bounded below, since the sequence H_t is required to be bounded.

The absence of arbitrage in this market follows from three factors: the date of the price increase is unpredictable; the price will be arbitrarily low even after the increase if ω is sufficiently large; and only bounded trading strategies are allowed. Recall the definition of an arbitrage opportunity as well as the fact that $V_t = V_t^*$ since $r = 0$. If $V_0 = 0$ and $\{H_t\}$ is such that $V_t \to V$ with $V(\omega) \geqslant 0$ for all ω, then

$$\textbf{(7.15)} \qquad -\sum_{t=1}^{\omega-1} (1/2)^t H_t + (\omega^2 + 2\omega)(1/2)^\omega H_\omega \geqslant 0, \text{ all } \omega \in \Omega$$

Now suppose for some integer k and some $\varepsilon > 0$ that

$$\textbf{(7.16)} \qquad \sum_{t=1}^{k} (1/2)^t H_t > \varepsilon$$

It follows easily by induction, using (7.15), that

$$(\omega^2 + 2\omega)(1/2)^\omega H_\omega > \varepsilon, \quad \text{all } \omega > k$$

But this cannot be true because $\{H_t\}$ is bounded, so there does not exist any integer k and $\varepsilon > 0$ such that (7.16) is true. In other words, it must be that

(7.17) $$\sum_{t=1}^{k} (1/2)^t H_t \leqslant 0, \quad \text{all } k \geqslant 1$$

Taking $\omega = 1$ in (7.15) and $k = 1$ in (7.17), it follows that $H_1 = 0$. More generally, if $H_1 = H_2 = \ldots = H_{k-1} = 0$, then (7.15) and (7.17) imply $H_k = 0$. Thus, by induction, our candidate for an arbitrage opportunity satisfies $H_t = 0$ for all t. There cannot exist any arbitrage opportunities.

To show there cannot exist any risk neutral probability measures Q, consider what the risk neutral conditional probabilities must be. Let q_{t-1} denote the risk neutral conditional probability of an "up" move from time $t-1$ to time t. The corresponding conditional expectation of ΔS_t must be zero, that is,

$$q_{t-1}(t^2 + 2t)(1/2)^t + (1 - q_{t-1})[-(1/2)^t] = 0. \quad \text{all } t \geqslant 1$$

This implies

(7.18) $$q_{t-1} = (t+1)^{-2}, \quad \text{all } t \geqslant 1$$

Another induction argument shows that the unconditional risk neutral probability $Q(\omega \geqslant t)$ must therefore equal $(t+1)/(2t)$, which, it should be noted, converges to 1/2 as $t \to \infty$. But for this to be a valid probability measure, it must be that

$$\lim_{t \to \infty} Q(\omega \geqslant t) = 0$$

whereas here this limit equals 1/2. Thus there are no probability measures under which the price process S is a martingale.

While example 7.2 is a discouraging result, it is not the end of the world. In the first place, while some might argue that the security model is a reasonable one from the economic point of view, others would argue that it is not. Modest variations on the definition of an arbitrage opportunity could lead to the identification of undesirable trading strategies.[2] These variations can become very technical, so they will not be pursued here. In the second place, we had to work quite hard to come up with an example like this. Such examples are rare. In practice, risk neutral probability measures invariably exist for realistic models of securities markets, even if there is an infinite number of trading periods.

Exercise 7.3 Show in detail that (7.15) and (7.16) imply $(\omega^2 + 2\omega)\ (1/2)^\omega H_\omega > \varepsilon$ for all $\omega > k$.

Exercise 7.4 Show in detail that (7.15) and (7.17) imply $H_t = 0$ for all $t = 0$.

Exercise 7.5 Show in detail that (7.18) implies $Q(\omega \geqslant t) = (t+1)/(2t)$.

NOTES

1 This result and much of the subsequent development is based on Dalang, Morton and Willinger (1990). See also Schachermayer (1992).

2 See Schachermayer (1994) for a comprehensive, advanced study of this subject.

Appendix: Linear Programming

Linear programs are optimization problems where there is a linear objective function in one or more variables and there is also a number of linear constraints. Linear programs have many useful applications, such as the scheduling of transportation systems, the control of production and inventory systems, and the optimal allocation of scarce resources. The study of linear programs is a primary topic in the subject of operations research. Many useful books have been written on the subject, ranging from the classical book by Dantzig (1998), through the excellent treatments by Murty (1976) and Chvatal (1980), and up to the recent texts by Sierksma (1996) and Vanderbei (1998). The purpose of this appendix is to briefly summarize the aspects of linear programming that are utilized for the theory of securities markets in this book.

A typical linear program is of the form

$$\text{(A.1)} \qquad \begin{aligned} &\textit{minimize} && c'X \\ &\textit{subject to} && AX \geqslant b \\ & && X \geqslant 0 \end{aligned}$$

where the decision variables comprise the non-negative column vector $X \in \mathbb{R}^n$ and the data consist of the column vector $c \in \mathbb{R}^n$, the column vector $b \in \mathbb{R}^m$, and the $m \times n$ matrix $A = (a_{ij})$. The feasible region, that is $\{X \in \mathbb{R}^n : AX \geqslant b,\ X \geqslant 0\}$, is either empty or a convex subset of $\mathbb{R}^n$, being the intersection of $m \times n$ half-spaces. Moreover, the feasible region, if it is non-empty, will have a number of *corners*, *vertices*, or *extreme points*, that is, feasible points which cannot be expressed as convex combinations of other feasible points. Since the objective function is linear, if an optimal solution exists, then at least one extreme point will be optimal.

Linear programs are solved by using one of two kinds of algorithms. The *simplex algorithm* and its variations proceed from one extreme point to another in a systematic fashion until the optimal extreme point is reached. The *ellipsoid method* ignores the extreme points but rather focuses on solving a certain system of inequalities; it enjoys the theoretical property of having a computational time that is bounded above by a fixed polynomial in the size of the data. Both kinds of algorithms are readily available in the form of commercial computer code.

There are several variations to the form in (A.1): the objective function can be maximized, some of the m constraints can be equalities and/or reverse inequalities, and some of the variables can be unrestricted in sign. However, all such variations can be transformed to the *standard form* (A.1). For instance, maximizing $c'x$ is the same as minimizing $-c'x$, and the constraint $\sum_j a_{ij}X_j \leqslant b_i$ is the same as $\sum_j(-a_{ij})X_j \geqslant -b_i$. Also, the constraint $\sum_j a_{ij}X_j = b_i$ is the same as the pair of constraints $\sum_i a_{ij}X_j \leqslant b_i$ and $\sum_i a_{ij}X_j \geqslant b_i$. Finally, if the variable X_j is unrestricted in sign, then it is equivalent to replace it everywhere in the problem by the difference, say $U_j - V_j$, of two non-negative variables.

Every linear program (LP) is paired with its *dual*. For instance, the dual of the LP in standard form (A.1) is

$$\text{(A.2)} \qquad \begin{aligned} &\text{maximize} && Y'b \\ &\text{subject to} && A'Y \leqslant c \\ & && Y \geqslant 0 \end{aligned}$$

where $Y \in \mathbb{R}^m$ is a column vector of variables. Using the transformations in the preceding paragraph, it is straightforward to show that the dual of (A.2) is the original LP (A.1) (see exercise A.1). This principle is true for any of the variations discussed in the preceding paragraph; in general, with the convention of calling the original LP the *primal*, one says that 'the dual of the dual is the primal.'

It is worthwhile spending a minute considering the duals of the different variations discussed above. First note that the m primal constraints are in one-to-one correspondence with the dual variables. Similarly, the n primal variables are in one-to-one correspondence with the dual constraints. To establish the dual of an arbitrary LP, one begins by making sure that if the primal objective function is to be minimized, then there are no "less than or equal to" (that is, "$\leqslant$") constraints. Similarly, if the primal objective function is to be maximized, then there must not be any "greater than or equal to" (that is, "$\geqslant$") constraints.

Next, as in going between (A.1) and (A.2), if the primal is "maximization" then the dual is "minimization," and vice versa. The data in the primal objective function becomes the data in the dual right hand side, and vice versa. If the data in the left hand side of the primal constraints is organized in the matrix A, then the corresponding dual matrix becomes the transpose A' (keep in mind that $(AX)' = X'A'$). Moreover, if a primal variable is required to be non-negative, then the corresponding dual constraint is an inequality ("$\leqslant$" if the dual objective function is to be maximized; "$\geqslant$" otherwise). Furthermore, if a primal constraint is an inequality, then the corresponding dual variable is required to be non-negative.

Now for the variations. If a primal variable is unrestricted in sign, then the corresponding dual constraint is an equality (see exercise A.2). And if a primal constraint is an equality, then the corresponding dual variable is unrestricted in sign (see exercise A.3).

Every linear program can be classified into one of three categories, as indicated in the following:

(A.3) Given an arbitrary linear program, either

(a) it has a finite optimal solution,
(b) the optimal solution is unbounded, or
(c) the feasible region is empty.

For example, minimizing X_1 with respect to the non-negative variables X_1 and X_2 is an example of (a), whereas changing this to a maximization problem gives an example of (b). Adding the constraint $X_1 + X_2 \leqslant -7$ gives an example of (c).

There is a logical relationship between the category for a linear program and the category of its dual. Suppose X is feasible for (A.1) and Y is feasible for (A.2). Multiplying the primal constraints on the left by the row vector Y' gives $Y'AX \geqslant Y'b$. Meanwhile, the dual constraints are the same as $Y'A \leqslant c'$, so multiplying this on the right by the column vector X gives $Y'AX \leqslant c'X$. Combining these inequalities give $c'X \geqslant Y'b$. Since X and Y are arbitrary feasible points, some logical conclusions are:

(A.4) If a linear program is in category (A.3a), then its dual cannot be in category (A.3b). Moreover, if a linear program is in category (A.3b), then its dual must be in category (A.3c).

If a linear program is in category (A.3a), then can its dual be in (A.3c)? The answer is no:

(A.5) If a linear program is in category (A.3a), then so is its dual.

Before proving this result, it should be mentioned that it is possible for a linear program as well as its dual to both be in category (A.3c) (see exercise A.4). Thus if a linear program is in category (A.3c), it follows that its dual is in either category (A.3b) or (A.3c), but it cannot be in (A.3a).

In order to verify (A.5), use will be made of Farkas's Lemma, which is derived from the separating hyperplane theorem. Although this lemma was already stated for exercise 1.11, it will be repeated here for convenience:

(A.6) Given the matrix D and the row vector d, either there exists a column vector v such that $Dv \leqslant 0$ and the scalar dv is strictly positive or there exists a non- negative row vector w such that $wD = d$, but not both.

To apply this lemma, take for D the matrix consisting of $n + m$ rows and $m + 1$ columns with (A', c) comprising the first n rows and $(-I, 0)$ comprising the last m rows, where I is the $m \times m$ identity matrix. Set d equal to the $m + 1$-dimensional row vector $(0, \ldots, 0, -1)$. It is left to the reader (exercise A.5) to show that (A.2) is feasible if and only if there exists some v such that $Dv \leqslant 0$ and dv is strictly positive.

We now prove (A.5) by assuming linear program (A.1) is in category (A.3a) while LP (A.2) is in category (A.3c) and then deriving a contradiction.

By the preceding paragraph and (A.6), if LP (A.2) is in category (A.3c), then there exists a non-negative row vector $w = (w_1, w_2)$ such that $wD = d$, where w_1 has n components and w_2 has m components. Writing this equation explicitly yields two equations, one of which is $w_1 c = -1$. The other equation is $w_1 A' - w_2 = 0$, which is really the same as $Aw_1' \geqslant 0$. So the infeasibility of (A.2) implies the existence of a w_1 such that $Aw_1' \geqslant 0$ and $w_1 c = -1$.

Now let X be a feasible solution to the primal LP (A.1) and consider the point $X + \lambda w_1$, where λ is any non-negative scalar. Clearly $X + \lambda w_1$ is non-negative for all non-negative values of λ. Moreover, $A(X + \lambda w_1) \geqslant b$, so, in fact, $X + \lambda w_1$ is primal feasible for all non-negative λ. The objective function is $c(X + \lambda w_1) = cX - \lambda$, so letting λ increase to infinity it is apparent that the objective function decreases to minus infinity. This contradicts our assumption that the primal LP (A.1) is in category (A.3a).

It turns out that the conclusion of (A.5) can be strengthened. The proof of the following result is similar to the proof of (A.5), and so it will be left to the reader (see exercise A.6).

(A.7) The duality theorem: If a linear program is in category (A.3a), then so is its dual, and the respective optimal values of the objective functions are equal to each other.

Exercise A.1 Use the transformations 'maximizing cX is the same as minimizing $-cX$' and so forth to prove that (A.1) is the dual of the dual of (A.1).

Exercise A.2 Use the aforementioned transformations and the duality between (A.1) and (A.2) to show that if a primal variable is unrestricted in sign, then the corresponding dual constraint is an equality.

Exercise A.3 Use the aforementioned transformations and the duality between (A.1) and (A.2) to show that if a primal constraint is an equality, then the corresponding dual variable is unrestricted in sign.

Exercise A.4 Show that neither the linear program

$$\begin{aligned} \text{minimize} \quad & X_1 - 2X_2 \\ \text{subject to} \quad & X_1 - X_2 \geqslant -1 \\ & -X_1 + X_2 \geqslant 2 \\ & X_1 \geqslant 0,\ X_2 \geqslant 0 \end{aligned}$$

nor its dual have feasible solutions.

Exercise A.5 With D and d defined as in the proof of (A.5), show that there is no v such that $Dv \leqslant 0$ and that dv is strictly positive if and only if LP (A.2) is infeasible. In other words, show there exists some v such that $Dv \leqslant 0$ and dv is strictly positive if and only if LP (A.2) is feasible.

Exercise A.6 Prove the duality theorem (A.7). Hint: if it is false, then there cannot be any solutions to $AX \geqslant b$, $Y'A \leqslant c$, $X \geqslant 0$, $Y \geqslant 0$, and $c'X \leqslant Y'b$.

Bibliography

Bartle, Robert G. and Sherbert, Donald R. 1992: *Introduction to Real Analysis*, Wiley, New York.

Baxter, Martin and Rennie, Andrew 1996: *Financial Calculus*, Cambridge University Press, Cambridge.

Bazaraa, M. S. 1993: *Nonlinear Programming*, Wiley, New York.

Berberian, Sterling K. 1994: *A First Course in Real Analysis*, Springer-Verlag, New York.

Bertsekas, Dimitri P. 1976: *Dynamic Programming and Stochastic Control*, Academic Press, New York.

Bingham, N. H. and Kiesel, Rüdiger 1998: *Risk Neutral Valuation*, Springer-Verlag, New York.

Bjork, Tomas 1998: *Arbitrage Theory in Continuous Time*, Oxford University Press, Oxford.

Browder, Andrew, Halmos, P. R. and Axler, S. 1996: *Mathematical Analysis: An Introduction*, Springer Verlag, New York.

Brown, William C. 1991: *Matrices and Vector Spaces*, Marcel Dekker, New York.

Chiang, Alpha C. 1974: *Fundamental Methods of Mathematical Economics*, McGraw Hill, New York.

Chvatal, V. 1980: *Linear Programming*, W.H. Freeman & Co., New York.

Cinlar, Erhan, 1975: *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, NJ.

Cox, John C. and Rubenstein, Mark 1985: *Options Markets*, Prentice-Hall, Englewood Cliffs, NJ.

Dalang, R., Morton, A. and Willinger, W., 1990: "Equivalent martingale measures and no-arbitrage in stochastic securities market models", *Stochastics and Stochastic Reports*, 29, pp. 185–201.

Dantzig, G. B. 1998: *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ.

Denardo, Eric V. 1982: *Dynamic Programming: Models and Applications*, Prentice-Hall, Englewood Cliffs, NJ.

Dixit, Avinash K. 1990: *Optimization in Economic Theory*, Oxford University Press, Oxford.

Dixit, Avinash K. and Pindyck, Robert S. 1994: *Investment Under Uncertainty*, Princeton University Press, Princeton, NJ.

Doob, J. L. 1953: *Stochastic Processes*, Wiley, New York.

Dothan, Michael U. 1990: *Prices in Financial Markets*, Oxford University Press, Oxford.

Duffie, Darrell 1988: *Security Markets: Stochastic Models*, Academic Press, New York.

—— 1992: *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, NJ.

Durrett, Richard 1991: *Probability: Theory and Examples*, Wadsworth & Brooks, Pacific Grove, CA.

Eatwell, John, Milgate, Murray and Newman, Peter 1989: *The New Palgrave: Finance*, W. W. Norton, New York.

Elliott, Robert J. and Kopp, P. Ekkehard 1998: *Mathematics of Financial Markets*, Springer-Verlag, New York.

Feller, William 1968: *An Introduction to Probability Theory and Its Applications*, Vol.1, 3rd edn., Wiley, New York.

—— 1971: *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edn. Wiley, New York.

Gantmacher, F. R. 1959: *The Theory of Matrices*, 2 vols, Chelsea, New York.

Harrison, J. Michael and Pliska, Stanley, R. 1981: "Martingales and stochastic integrals in the theory of continuous trading," *Stochastic Processes and Their Applications*, 11, pp. 215–60.

Hayhurst, George 1987: *Mathematical Programming Applications*, Macmillan, New York.

Hoel, P., Port, S. and Stone, C. 1972: *Introduction to Stochastic Processes*, Houghton Mifflin, Boston.

Huang, Chi-fu and Litzenberger, Robert H. 1988: *Foundations for Financial Economics*, North-Holland, New York.

Hull, John 1997: *Options, Futures, and Other Derivative Securities*, Prentice Hall, Englewood Cliffs, NJ.

Ingersoll, Jr., Jonathan E. 1987: *Theory of Financial Decision Making*, Rowman & Littlefield, Totowa, NJ.

Jarrow, Robert A. 1988: *Finance Theory*, Prentice-Hall, Englewood Cliffs, NJ.

Jarrow, Robert and Turnbull, Stuart 1996: *Derivative Securities*, South-Western College Publishing, Cincinnati.

Jeter, Melvyn W. 1986: *Mathematical Programming: An Introduction to Optimization*, Marcel Dekker, New York.

Karatzas, Ioannis and Shreve, Steven E. 1998: *Methods of Mathematical Finance*, Springer-Verlag, New York.

Karlin, Samuel and Taylor, Howard M. 1975: *A First Course in Stochastic Processes*, Academic Press, New York.

—— 1981: *A Second Course in Stochastic Processes*, Academic Press, New York.

Karr, Alan F. 1993: *Probability*, Springer-Verlag, New York.

Kijima, Masaaki 1997: *Markov Processes for Stochastic Modeling*, Chapman & Hall, London.

Klein, E. 1973: *Mathematical Methods in Theoretical Economics*, Academic Press, New York.

Korn, Ralf 1997: *Optimal Portfolios*, World Scientific Publishing Co., Singapore.

Lamberton, Damien and Lapeyre, Bernard 1996: *Introduction to Stochastic Calculus Applied to Finance*, Chapman & Hall, London.

Lay, David C. and Guardino, Karen 1997: *Linear Algebra and Its Applications*, Addison-Wesley, Harlow, Essex.

Luenberger, David G. 1998: *Investment Science*, Oxford University Press, Oxford.

Markowitz, Harry 1990: *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, Blackwell, Oxford.

Merton, Robert C. 1990: *Continuous-Time Finance*, Blackwell, Oxford.

Mikosch, Thomas 1998: *Elementary Stochastic Calculus with Finance in View*, World Scientific Publishing Co., Singapore.

Mikusinski, Jan and Mikusinski, Piotr 1993: *An Introduction to Real Analysis: From Number to Integral*, Wiley, New York.

Murty, Katta G. 1976: *Linear and Combinatorial Programming*, Wiley, New York.

Musiela, Marek and Rutkowski, Marek 1997: *Martingale Methods in Financial Modeling*, Springer-Verlag, New York.

Neveu, J. 1975: *Discrete-parameter Martingales*, North-Holland, Amsterdam.

Norris, J. 1998: *Markov Chains*, Cambridge University Press, Cambridge.

Olkin, Ingram, Gleser, Leon J. and Derman, Cyrus 1980: *Probability Models and Applications*, Macmillan, New York.

Ostaszewski, Adam 1993: *Mathematics in Economics: Models and Methods*, Blackwell, Oxford.

Panjer, Harry et al. 1998: *Financial Economics, With Applications to Investments, Insurance and Pensions*, The Actuarial Foundation, Schaumburg, Illinois.

Pitman, Jim 1993: *Probability*, Springer-Verlag, New York.

Pliska, Stanley R. 1982: "A discrete time stochastic decision model," in *Advances in Filtering and Optimal Stochastic Control*, (edited by W. H. Fleming and L. G. Gorostiza; Lecture Notes in Control and Information Sciences, 42), Springer-Verlag, New York, pp. 290–304.

—— 1986: "A stochastic calculus model of continuous trading: optimal portfolios," *Math. Operations Research*, 11, pp. 371–84.

Puterman, Martin L. 1994: *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley, New York.

Rebonato, Riccardo 1998: *Interest Rate Option Models: Understanding, Analyzing and Using Models for Exotic Interest-Rate Options*, John Wiley, Chichester.

Revuz, D. 1984: *Markov Chains*, North-Holland, Amsterdam.

Riess, R. Dean, Johnson, Lee W. and Arnold, Thomas A. 1997: *Introduction to Linear Algebra*, Addison-Wesley, Harlow, Essex.

Rockafellar, R. Tyrrell 1997: *Convex Analysis*, Princeton University Press, Princeton, NJ.

Roman, Steven 1992: *Advanced Linear Algebra*, Springer-Verlag, New York.

Ross, Sheldon M. 1995: *Stochastic Processes*, John Wiley, Chichester.

—— 1997a: *A First Course in Probability*, Prentice-Hall, Englewood Cliffs, NJ.

—— 1997b: *Introduction to Probability Models*, Academic Press, New York.

Schachermayer, W., 1992: "A Hilbert space proof of the fundamental theorem of asset pricing in finite discrete time", *Insurance: Mathematics and Economics*, 11, pp. 1–9.

—— 1994: "Martingale measures for discrete-time processes with infinite horizon", *Mathematical Finance*, 4, pp. 25–56.

Sierksma, Gerard 1996: *Linear and Integer Programming: Theory and Practice*, Marcel Dekker, New York.

Taylor, Howard M. and Karlin, Samuel 1984: *An Introduction to Stochastic Modeling*, Academic Press, New York.

Vanderbei, Robert J. 1998: *Linear Programming: Foundations and Extensions*, Kluwer Academic Press, Kingston-upon-Thomes, Surrey.

Whittle, Peter 1982: *Optimization Over Time: Dynamic Programming and Stochastic Control*, Vol. 1, Wiley, New York.

—— 1983: *Optimization Over Time: Dynamic Programming and Stochastic Control*, Vol. 2, Wiley, New York.

Williams, David 1991: *Probability with Martingales*, Cambridge University Press, Cambridge.

Wilmott, Paul, Dewynne, Jeff and Howison, Sam 1993: *Option Pricing: Mathematical Models and Computation*, Oxford Financial Press, Oxford.

Index