Mathematics in Finance

Robert Almgren
University of Chicago

MAA Short Course
San Antonio, Texas
January 11-12, 1999

Contents

Chapter 1

Completeness and Arbitrage

Let's begin with some of the main ideas and methods of mathematical finance. These may seem a little abstract at first, but the ideas of *arbitrage* (more often, of its absence) and of a *complete market* underly everything else. The existence of these simple models and solutions has a strong influence on what kinds of models we are willing to make of the messy real world, so I would rather get these ideas on the table first, before we start talking about real stocks, bonds, and options.

1.1 One-period model

In the first model we shall consider, there are two times. There is now, $t = 0$, at which we know everything about the world. There is a future time $t = T$, at which we do not know exactly what will happen. We make investment decisions now which will, in general, lead to *uncertain* outcomes at time T.

1.1.1 Future states

The classic theory relies on the following formulation: We make a list of the possible future states of the world. It is important for this theory that this list is *finite*; if the number of possible future states is M, then we can just label states by numbers 1 to M. Let us call the set of states $\Omega = \{1, \dots, M\}$.

Here are some examples:

- A coin flip has two states: heads or tails. A lottery ticket either wins one of a few specified payoffs or nothing.

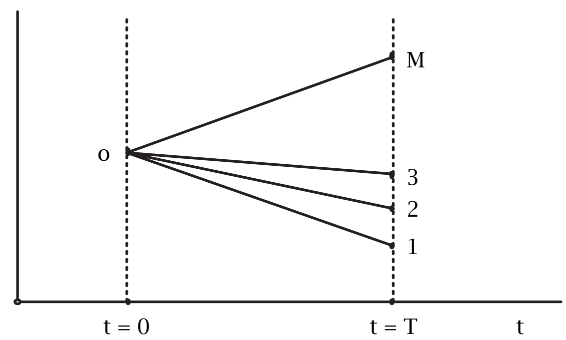

Figure 1.1: M states at future time $t = T$.

- Either the Republican or the Democratic candidate will win election in my district.

- If I am buying earthquake insurance, then I distinguish two states: either there is an earthquake (between now and time T, the expiration date of my policy) that knocks my house down, or there is not. Of course, the insurance company and I have to agree on what constitutes an insurable event. Actually, the policy might also cover partial damage, so the number of states may be greater than two, but the two-state model might be a reasonable one to start with.

- Weather: I could measure the number of heating-degree days in a specific month, and divide it into bins to make the state space finite. Oil prices move up and down.

- Most relevantly for our purposes, we might construct a model for the price of a specified stock, in which at a specified future time, the price can take only one of a few finite values. Since stock prices vary nearly continuously, this would clearly be an approximation, but it will turn out to be a useful one. Here the "event" is simply the motion of the price itself; we are not claiming that it depends on any particular outside influence.

One thing we are explicitly *not* going to include in the theory at this point is any opinion about the *probability* with which these different events will occur. All we need is a finite list of *possible* events. (As Mr. Spock once said, "This is not about probabilities, Lieutenant. We must be logical.")

It might be more accurate to say that we care only whether probabilities are zero or nonzero, but we don't care what their values are. We implicitly assign zero probability to all events that are not in our list Ω (the coin might land on its edge).

A *random variable* is a function $\Omega \to \mathbb{R}$ (real numbers). In other words, it is a vector $(f(1), \ldots, f(M))$, where $f(j)$ is the value this variable will take if state j happens. A random variable can be represented as a (row) vector in the Euclidean space $\mathbb{R}^M$.

1.1.2 Endowment and consumption

In economics jargon, we measure our wealth in terms of a rather abstract quantity called a *consumption good*. In classical theory, consumption is not the same as money; money is a social invention, which is useful insofar as it can be exchanged for the consumption good.

For example, if I acquire a lot of money, I can buy a Porsche. My consumption is not the car itself, but the experience of driving around in it, cornering as if on rails and receiving the envy of my neighbors. It is my decision whether the consumption I sacrificed to achieve this state, for example, the time I spent working in the office, is worth the result.

For our purposes, we can equate consumption with money. That is, we are considering only a restricted set of the possible measures of success. For an individual, it might well be the case that applying these theories is more trouble that it is worth, indicating that important consumption quantities have been left out of the model. For large financial institutions, the relative cost is much less.

The reason we are given anything to consume is that we are given an *endowment*; in our model, this is an amount of money we receive from some outside source depending on time and on the state of the world. Let us consider the endowmwent in the examples above:

- A coin flip does not in itself give me or cost me any money. (I can enter into a contract with someone else, the value of which depends on the result of the flip.) A lottery ticket is similar: I have no exposure unless I choose to buy it.

- It is possible that I will have financial exposure to the outcome of an election. Perhaps I expect that a Republican win will drive the stock market upwards or downwards. Perhaps I expect a cushy job in a Democratic administration.

- I spend money to buy or build a house. If no earthquake happens I keep the house so I have no new endowment. If an earthquake knocks it down, then I have a huge negative endowment.

- Energy companies do have financial exposure to monthly average temperature, since consumption and consequently price depend on it. Airlines have huge exposure to oil prices, since they must buy large quantites.

- The stock motion is like the coin flip: all kinds of stocks go up and down all the time without concerning me. The only way I would have exposure to the future stock motion is if at $t = 0$ I decide to purchase some shares, but this is not considered an endowment in the classical sense.

Our future endowment is a random variable, which is not under our control. Our future consumption is also a random variable. Total endowment can be represented as the $(M + 1)$-vector (e_0, e), where the scalar e_0 is the endowment at the initial time, and e is the M-vector (random variable) representing endowment at the final time. Similarly, total consumption can be represented (c_0, c).

The purpose of financial markets is to provide mechanisms for making consumption different from endowment. For example,

- I can choose to bet on a coin flip, or to buy a lottery ticket. If I do, my consumption becomes very non-uniform.

- If I buy an insurance contract, I make my net consumption much less random (I pay a small amount whatever happens, but I am guaranteed to still have a place to live).

- Energy companies buy and sell contracts whose value depends on temperature, in order to eliminate the risk they are already exposed to. Airlines buy futures and options on fuel, to reduce or eliminate their risk.

- I may choose to buy a stock: in this case I generally am hoping that the price will rise and my consumption will, on average, be higher than if I had not bought.

Another favorite economics concept is the *utility function*, which measures the net value to me of a random consumption variable. This may be represented as a function $U(x)$ from $\mathbb{R}^M$ to $\mathbb{R}$; that is, it depends on the entire spectrum of possible outcomes.

For example, when I decide to purchase earthquake insurance, I am judging that my utility function is controlled by the *worst* possible outcome. Thus I am happy to protect myself against a downside risk, even if statistically (assigning probabilities to the events) my average gain is negative. Fortunately, the insurance company has a different utility function (and different endowment), so the trade is mutually beneficial.

Utility functions are generally *increasing*, meaning that more is always better, and *concave*, modeling to aversion to risk. Thus, if I have made a million dollars, I would still like to make more, but the second million will change my life less than did the first.

For our purposes, all we need to assume is that everyone in the market has an increasing utility function. The reason is that we will be considering the very special class of models and markets in which it is possible to completely *eliminate* risk. This is the theory that underlies the pricing of financial derivatives such as options, and is the part that gives the most classical and cleanest mathematics. Needless to say, this is a very special category, and very large and important areas of financial mathematics are devoted to the modeling and management of risk.

1.1.3 Securities and markets

A *security* is a contract that pays different amounts in different states of the world; that is, it is a random variable $d = (d(1), \dots, d(M)) \in \mathbb{R}^M$.

For example, a bet on a coin flip would be represented as $(1, -1)$. An earthquake insurance contract might be represented as $(0, P)$; it pays a large amount P in the event of an earthquake. A stock might be represented as, for example, $(90, 100, 110)$, corresponding to the three possible future prices.

We shall assume that every security can be purchased either *long*, meaning you hold a positive amount, or *short*, meaning that you hold a negative amount. That is, you can buy any real number θ of "shares," and then your payout is the random variable $(\theta d(1), \dots, \theta d(M))$. You control this outcome by selecting θ at $t = 0$.

Thus you can choose either heads or tails on a coin flip. You can either buy an insurance contract, or write one to your neighbor. You can buy a stock, or "short" it, meaning that you borrow shares, sell them, and are obliged to purchase back the shares at a later date; you make money if the price has dropped in the meantime. Clearly the assumption of free long or short investment is not sensible for every security, but for the main ones traded in financial markets it is reasonably good.

A *market* is a list of securities $d_1, \ldots, d_N$. The market may be represented by its $N \times M$ *payout matrix*

$$D = \begin{pmatrix} d_1(1) & \cdots & d_1(M) \\ \vdots & & \vdots \\ d_N(1) & \cdots & d_N(M) \end{pmatrix}$$

giving the amount that each security pays in each state of the world. The number of securities N may be larger than, smaller than, or equal to the number of possible states M.

A *portfolio* is a list of investments

$$\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_N \end{pmatrix}$$

where θ_j is the amount of security j you purchase at $t = 0$. The net payout yielded by portfolio theta is the matrix product $D^{\mathsf{T}}\theta$, where $^{\mathsf{T}}$ denotes transpose. The important point is that you must determine your investment θ *before* the state is revealed. Thus the net payout is a random variable. Choosing the portfolio is your means for controlling your consumption in the presence of uncertain endowment.

The *prices* of the securities are given by the N-vector

$$p = \begin{pmatrix} p_1 \\ \vdots \\ p_N \end{pmatrix}$$

where p_j is the money you have to invest to acquire one unit of security j. Thus the total cost of portfolio θ is the scalar $p^{\mathsf{T}}\theta$; this is not a random variable since p is known at the initial time.

For a given portfolio θ, your total consumption, meaning the amount of money you extract from the investment, is the $(M+1)$-vector $(e_0, e) + \bar{D}^{\mathsf{T}}\theta$, where the $N \times (M+1)$ augmented matrix

$$\bar{D} = \begin{pmatrix} -p & D \end{pmatrix} = \begin{pmatrix} -p_1 & d_1(1) & \cdots & d_1(M) \\ \vdots & \vdots & & \vdots \\ -p_N & d_N(1) & \cdots & d_N(M) \end{pmatrix}$$

contains prices and payouts together. This matrix completely describes the market. When these prices and payouts are posted, you are free to buy or sell as much of each security as you would like, in order to tailor your final consumption however you would like.

So far we have said nothing about how these prices are determined; they are offered to us and to every other participant by some mysterious marketplace "out there." The *entire* point of this course is that the simple principle of "no free lunch" (nonexistence of arbitrage, as defined below) imposes certain constraints on the prices. In particular,

1. We can write inequality conditions that must be satisfied by the current prices in terms of their final payouts (the lesser point).

2. Most importantly: If our market is "complete" (see below), then when we add a new security into the market, we can determine *exactly* what its price must be in terms of the prices and payouts of existing securities. The classic example is option pricing: if someone proposes a new contract whose value depends on the value of something that already exists in the market, then, using suitable mathematical models, we can determine what the price of the option *must* be.

That's really all that classical financial mathematics consists of. Once you are clear on the basic idea, the things that make it interesting are, first, the mathematical mechanisms for carrying out this computation in various complicated situations, second, the kind of models about the real world this simple theory forces us to believe, and, third, how the simple theories are not adequate and what we can do about it.

In some cases you may already have opinions about what the price should be in terms of the payout. For example, coin flips and lotteries are canonical examples of problems to which probability theory can be applied. You would have difficulty getting someone to pay you to bet on a fair coin; the price of lottery tickets isset by the state and you may have stron opinions about whether they are fair (but you may buy one anyway from time to time). Remember that this theory does not care at all about probabilities except as they are zero or nonzero; it is really only based on the nonexistence of guaranteed profit.

Let us now consider two specific kinds of securities, which will be our building blocks.

A *bond* is a riskless asset. We will define a unit amount of a bond to pay a unit amount in all states of the world. Thus its payout at time

T is the constant vector $(B_1, \ldots, B_M) = (1, \ldots, 1)$. We can purchase this bond now for a discounted price, where the discount factor depends on (or determines) the *interest rate*. If we imagine that interest accumulates continuously at rate r per unit time, then the discount factor is e^{-rT}, though $1 - rT$ or $(1 + rT)^{-1}$ are also reasonable models. Since the payout is unity, this is also the price $B_0 = e^{-rT}$.

Positive investment in a bond (going long) means loaning money to whoever "issued" the bond: you put out cash now, and you get back cash later. Negative investment (going short) means borrowing money. Our model says that you can do both of these at the *same* interest rate r.

A *stock*, in this model, just represents a generic risky asset. Let's call the current price S_0; the price at time T has M possible values $(S_1, \ldots, S_M)$.

Thus in this two-asset market, the total $2 \times (M + 1)$ payout matrix is

$$\bar{D} = \begin{pmatrix} -e^{-rT} & 1 & \cdots & 1 \\ -S_0 & S_1 & \cdots & S_M \end{pmatrix}$$

1.1.4 Arbitrage and pricing measures

Now we define the notion of a riskless profit. An *arbitrage strategy* is a portfolio $\theta \in \mathbb{R}^N$ so that $\bar{D}^{\mathrm{T}}\theta \ge 0$ and $\bar{D}^{\mathrm{T}}\theta \ne 0$. (In this one-period model that the words "portfolio" and "strategy" are interchangeable.)

(For vectors x and y, $x \ge y$ and $x > y$ are interpreted componentwise: $x_j \ge y_j$ or $x_j > y_j$ respectively for *every* j; $x \ne 0$ means that *some* component $x_j \ne 0$.)

The vector $\bar{D}^{\mathrm{T}}\theta$ has $M + 1$ components, one for the initial time and M for the final states. Depending on which component is positive, an arbitrage strategy can be either one of the following two things:

1. A strategy that does not cost anything to initiate ($p^{\mathrm{T}}\theta \le 0$), that cannot lose anything ($D^{\mathrm{T}}\theta \ge 0$ in all states), and that generates positive income in at least one possible state ($(D^{\mathrm{T}}\theta)_j > 0$ for some $j = 1, \ldots, M$). Or

2. A strategy that generates positive money when it is initiated ($p^{\mathrm{T}}\theta < 0$) and does not lose money no matter what happens ($D^{\mathrm{T}}\theta \ge 0$).

Now, if our market admitted an arbitrage strategy, then everyone in the world would rush to invest in it (this is where we use the increasing utility functions). No one would want the other side of those trades, since the opposite side would lose money with certainty, and noone can have a risk exposure for which that is a good thing. There would thus be a huge unbalanced demand from the market participants. Although we have not said anything about how prices are adjusted, any reasonable mechanism for making prices respond to demand would cause the prices to move so that the arbitrage opportunity disappeared (in chemistry, this is called Le Chatelier's principle).

A market is in *equilibrium* when this unbalanced demand does not exist, so the listed prices represent actual values at which other participants are willing to take the other sides of the trades. The fundamental principle of our pricing methods is that

Arbitrage strategies cannot exist in equilibrium.

The principle of absence of arbitrage is closely related to the efficiency of markets, and the ability to quickly pounce on any profit opportunities.[1]

Our next job is to see what this rather abstract statement implies for the relationship between the prices and the payout functions. This is provided by the

Theorem: For a given market consisting of a payout and price matrix $\bar{D}$, nonexistence of an arbitrage strategy is equivalent to the existence of a *state-price vector* $\psi > 0$ so that $p = D\psi$.

Proof: First, suppose that such a $\psi > 0$ exists. Then for any strategy θ, $p^{\mathrm{T}}\theta = \psi^{\mathrm{T}} D^{\mathrm{T}}\theta$. If all components of $D^{\mathrm{T}}\theta$ are zero, then $p^{\mathrm{T}}\theta = 0$ and θ is not an arbitrage (it doesn't cost anything but you don't get anything). On the other hand, if any component of $D^{\mathrm{T}}\theta$ is positive, then, since all its components must be ≥ 0 and since $\psi > 0$, $p^{\mathrm{T}}\theta$ must be > 0, so this θ is not an arbitrage (you can get something, but it costs you initially).

The converse is more subtle. In $\mathbb{R}^{M+1}$, consider the positive cone

$$K = \{ x \in \mathbb{R}^{M+1} \mid x \geq 0 \text{ and } x \neq 0 \}.$$

[1] The efficient market hypothesis is closely associated with the University of Chicago Department of Economics. It happened one day that a Chicago professor was walking down the street with some of his students. One of them looked at the sidewalk and exclaimed "Look, a $20 bill!" The professor sagely remarked, "Don't bother with it: if it were a real one, someone would already have picked it up."

The point is not that an occasional bill on the sidewalk is impossible, but you should not expect it to last for very long, especially if the sidewalk is crowded.

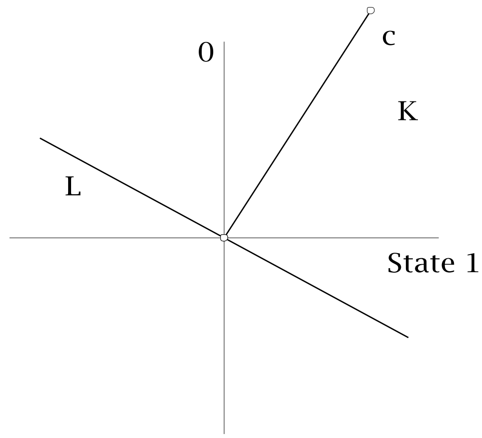

Figure 1.2: The arbitrage pricing theorem for $N = 1$, $M = 1$.

This is the set of payoffs (including initial cost) that constitute an arbitrage. Also consider the linear subspace

$$L = \left\{ x \in \mathbb{R}^{M+1} \mid x = \bar{D}^{\mathrm{T}}\theta \text{ for some } \theta \in \mathbb{R}^N \right\}.$$

This is the set of payoffs you can attain by some strategy; it has dimension $\mathrm{rank}(\bar{D})$. An intersection of K and L would be an arbitrage; if there do not exist arbitrage strategies, then K and L are disjoint.

Now, both K and L are convex sets, and L is closed and linear. Then the *separating hyperplane theorem* says that there exists a linear function F so that $F(x) = 0$ for all $x \in L$ and $F(y) > 0$ for all $y \in K$.

By the *Riesz representation theorem* (trivial in finite dimensions), the linear function can be represented as $F(x) = c^{\mathrm{T}}x$ for some vector $c \in \mathbb{R}^{M+1}$; c is a normal vector to L. We can write $c^{\mathrm{T}}x = ax_0 + b^{\mathrm{T}}\tilde{x}$, where $x = (x_0, \tilde{x})$. Since K contains the positive coordinate axes, each component $a > 0$, $b_j > 0$. Returning to the definition of L, this means that for every $\theta \in \mathbb{R}^N$, $-ap^{\mathrm{T}}\theta + b^{\mathrm{T}}D^{\mathrm{T}}\theta = 0$, or $(ap - Db)^{\mathrm{T}}\theta = 0$. Since $ap - Db$ is just a vector in $\mathbb{R}^N$, this can be true only if $ap = Db$, so $\psi = b/a$ is a state-price vector.

The state-price vector assigns a positive weight ψ_j to each state j. The theorem says that if there is no arbitrage, then each initial price p_i can be computed by "collapsing" the payout vector $(d(1), \ldots, d(M))$ against ψ, and we can use the *same* ψ for each security in our market. Note that if $\mathrm{rank}(\bar{D}) < M$, then the state-price vector is not necessarily unique.

Theorem: A sufficient condition for the state-price vector to be unique is that $\mathrm{rank}(\bar{D}) = M$.

Proof: As noted, $\mathrm{rank}(\bar{D})$ is the dimension of the subspace L. If this dimension is M, one less than the dimension of the space, then L has a unique normal vector c, the only candidate for the state-price vector.

Since $\bar{D}$ is an augmentation of D, $\mathrm{rank}(\bar{D}) \geq \mathrm{rank}(D)$ and we have the

Corollary: A sufficient condition for the state-price vector to be unique is that $\mathrm{rank}(D) = M$.

Since D is $N \times M$, its rank is at most the smaller of M and N. To have $\mathrm{rank}(D) = M$, there must be at least $N \geq M$ securities in the market, and their payoffs must be independent vectors. A market with $\mathrm{rank}(D) = M$ is called *complete*; in a complete market you can achieve any desired combination of payoffs in the different states, by suitable choice of investment at the initial time.

Note that if there are *more* securities than states, so $N > M$, then it would be possible for $\mathrm{rank}(\bar{D})$ to be $\geq M + 1$. In this case, the subspace L fills all of $\mathbb{R}^{M+1}$ and arbitrage is certainly possible.

If the number of possible future states is larger than the number of independent securities (in this one-period model) it is impossible for the market to be complete. So the restriction to finitely many possible future states is essential in this model. In order to let the price take arbitrarily many values, we shall also need to take many small time periods; it the relationship between the two continuum limits will set important restrictions on the nature of our model.

1.1.5 Risk-neutral pricing

In order to interpret the significance of the components of ψ, let us suppose that one element in our market is a bond B with discount factor $B_0 = e^{-rT}$. The above representation says that $B_0 = B_1\psi_1 + \cdots B_M\psi_M = \psi_1 + \cdots + \psi_M$. Thus if we define

$$q = e^{rT}\,\psi,$$

then each $q_j > 0$ and their sum is one. We may thus interpret each q_j as a "probability" that our model leads us to assign to state j; then absence of arbitrage means that we can compute the price of *any* security in the market by the formula

$$p = e^{-rT}\,\mathbb{E}_Q[d] \equiv e^{-rT}\big(q_1 d(1) + \cdots + q_M d(M)\big).$$

Here the notation $\mathbb{E}_Q$ means "expected value using the numbers q as probabilities." We are using the language of probability theory, but let us emphasize that the q_j do *not* represent any sort of estimate of the probability of anything happening in the real world. They are purely mathematical constructions; their values are forced on us by the relationships between the prices and the payouts of the securities in the market, and the assumption of nonexistence of arbitrage.

This result is so important and so fundamental that we will state it again in words:

> *If the market does not admit arbitrage, then a purely artificial probability measure can be constructed so that the price of any security traded in the market is equal to the expectation of its future value in that measure, discounted at the same rate as a risk-free bond. If the market is complete, this measure is unique.*

This probability measure is often called the "risk-free measure."

1.1.6 Valuing a single derivative

Let us now return to the simple model we have discussed above, in which the market consists of two assets. The first is a bond, with present value $B_0 = e^{-rT}$ and sure payoff 1 in every state. The second asset is a "stock" (it does not really matter what it is) with present value S_0 and uncertain future value.

We shall suppose that only two future states are possible: the stock price moves to value S_1 or to value S_2, with $S_1 < S_2$. We repeat that these two states are not consequences of any external event in the world; they simply represent our uncertainty about the future motion of the stock price.

Now the augmented price matrix is

$$\bar{D} = \begin{pmatrix} -e^{-rT} & -S_0 \\ 1 & S_1 \\ 1 & S_2 \end{pmatrix}$$

in which the first column is the bond and the second column is the stock. The first row represents the initial time; the second and third rows are the two possible future states.

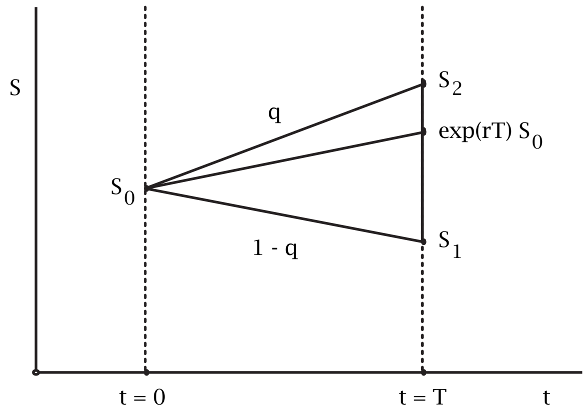

Figure 1.3: The 2-state stock motion model.

Since $S_1 \neq S_2$, $\bar{D}$ has rank 2. Thus the linear set L in the arbitrage theorem is a plane. It has a unique normal vector (a, b_1, b_2), and so the risk-neutral measure is unique if it exists.

Denoting $q_1 = 1 - q$, $q_2 = q$ (Figure 1.3), we need only solve

$$S_0 \;=\; e^{-rT}\Big(qS_2 + (1-q)S_1\Big)$$

which gives

$$q \;=\; \frac{e^{rT}S_0 - S_1}{S_2 - S_1}, \qquad 1 - q \;=\; \frac{S_2 - e^{rT}S_0}{S_2 - S_1}. \tag{1.1}$$

These are the normalized state-price vector, as long as $0 < q < 1$. By inspection, this will be the case if

$$S_1 \;<\; e^{rT}S_0 \;<\; S_2 \tag{1.2}$$

This constraint is easily understood in direct financial terms.

The theorem says that if $(q, 1-q)$ do not constitute a legitimate pricing measure, then there exists an arbitrage portfolio. Suppose, for example, that $e^{rT}S_0 \le S_1$. Then do the following investment: At $t = 0$, borrow S_0 cash, and use that money to purchase one share of stock, so you don't invest any of your own money. At $t = T$, you have to repay $S_0 e^{rT}$. Get that money by selling the stock, which yields either S_1 or S_2. In the first case you definitely don't lose, and in the second case you definitely make something since $S_2 > S_1$.

Conversely, if $S_1 < S_2 \le e^{rT} S_0$ then interest rates are so high that you should short the stock and loan out the money. The only way that neither of these strategies will not work is if the future stock value can be either more or less than the amount earned by investing its current value at the risk-free interest rate.

Now let's make it interesting. Suppose our market contains a stock S and a bond B, all of whose initial prices and possible payouts are specified so that there are no arbitrage possibilities. Now let's add an additional security V into the market.

In general, if we add something else, we increase the number of states. For example, if we considered a second stock, then that stock could in general move up or down independently of the first one, so we would have at least four possible states in the extended model.

We will suppose that the new security is a *derivative*, meaning that its value at the future time *depends* on the value of the *underlying asset*, in this case our original stock. There is an explicit function $\Lambda(S)$, the *payout function*, which gives the value of V in terms of the value of S at time T. Thus $V_1 = \Lambda(S_1)$ and $V_2 = \Lambda(S_2)$. The derivative security does not change its value independently of the underlying value (its value may depend on other parameters such as the interest rate).

An example (we will discuss this more thoroughly in the next section) would be a *call option* with expiration date T. A call option gives you the right, but not the obligation, to purchase the asset at the specified date for a prearranged value K, the *strike price*. If the asset is at that time trading in the market for a price greater than K, you make a net profit $S - K$; if it is trading for a price less than K the option is worthless. Thus $\Lambda(S) = \max\{S - K, 0\}$. (We are considering *European* options, which can be exercised only on a single future date.)

We now have $N = 3$ securities but still $M = 2$ future states. The 3×3 augmented payout matrix is

$$\bar{D} = \begin{pmatrix} -e^{-rT} & -S_0 & -V_0 \\ 1 & S_1 & V_1 \\ 1 & S_2 & V_2 \end{pmatrix}$$

in which the only unknown quantity is V_0.

But we can immediately determine what V_0 *must* be in terms of the other prices in the problem. Recall that for nonexistence of arbitrage opportunities, we need $\operatorname{rank}(\bar{D}) \le M$. With $M = 2$, $N = 3$, $\bar{D}$ must be degenerate.

The risk-neutral pricing formula gives us a simple way to express the algebra. Every security in the market must have a current price equal to the discounted risk-neutral expection of its future value, so in particular,

$$\begin{aligned} V_0 &= e^{-rT}\Big(qV_2 + (1-q)V_1 \Big) \\ &= \frac{S_0 - e^{-rT}S_1}{S_2 - S_1}\, V_2 + \frac{e^{-rT}S_2 - S_0}{S_2 - S_1}\, V_1. \end{aligned} \tag{1.3}$$

A different way to express this result also gives insight. Suppose we form a portfolio Π consisting of b units of the bond and Δ units of the stock (use of the Greek Delta is confusing but overwhelmingly customary). The present value of this portfolio is

$$\Pi_0 = b\, e^{-rT} + \Delta S_0.$$

The future value of this portfolio takes two different values depending on which way the stock price moves:

$$\begin{aligned} \Pi_1 &= b + \Delta S_1 \\ \Pi_2 &= b + \Delta S_2. \end{aligned}$$

Now we choose b and Δ so that the payoff of portfolio P is exactly the same as the payoff of the option: $\Pi_1 = V_1$ and $\Pi_2 = V_2$. In the jargon, we construct a portfolio that *replicates* the option. Since there are two states and two free parameters, we can do this uniquely (as long as $S_1 \neq S_2$), to obtain

$$\Delta = \frac{V_2 - V_1}{S_2 - S_1}, \qquad b = \frac{S_2V_1 - S_1V_2}{S_2 - S_1}. \tag{1.4}$$

It is easy to verify that then the present value of the portfolio $\Pi_0 = V_0$ as given in (1.3). Note that you have to solve the entire tree to find Δ at the starting time.

Suppose the option were being bought in the market for a price V_0' *greater* than this V_0. Then you could sell the option to someone, collecting cash V_0'. By selling the option, you incur risk; for example, if you sell a call option and the stock price has risen dramatically by time $t = T$, you will be obligated to deliver an expensive asset for a low price.

But in this model, you can perfectly *hedge* your risk. You buy Δ shares of the underlying, and invest whatever cash is left over at the going interest rate. If V increases with S (as for a call option, say), then you purchase a positive amount of stock, and the rise in value of the stock is exactly enough to cover your increased liability. You incur no risk, and walk away with a net profit. Because a lot of people will be trying to do this if $V_0' > V_0$, the market price will be quickly driven down to V_0. The quantity Δ is called the *hedge ratio*: the amount of stock you must hold per option you have sold to be risk-free.

This all sounds very clean and neat. But let's review some of the assumptions that went into the model, and were essential for the results:

- *You can buy or sell every asset in the market for the same price.* In reality, there are always transaction costs: brokerage fees, bid/ask spreads, *etc.*

- *You can buy arbitrary positive or negative amounts of every asset.* In reality, short stock selling does not work exactly the same as long purchasing: for example, there may be margin requirements on a short sale. Also, large transactions in the stock in order to cover options may move the stock price; we have assumed the stock moves completely independently. There are other strange effects: for example, the Chicago options exchange closes 15 minutes *after* the New York exchanges on which the underlying stocks are traded, and this theory says nothing about the price an option should have when the stock is not freely tradeable.

- Most importantly, in a discrete-time model, *you must specify the possible values to which the price can move in the next time period.* In the continuous-time limit, this corresponding to making a choice for the *volatility*. But volatility does not have an unambiguous value, and different people have different opinions. In practice, choosing an option price is equivalent to choosing a value for the volatility.

1.2 Binomial trees

The above model can be extended to multiple periods. For simplicity, we shall consider the model of the last section, containing only a bond, a stock, and eventually a derivative asset depending on the stock.

Let us suppose that time T is still the final horizon of our model, but let us now divide that time into N sub-times $t_0, t_1, \ldots, t_N$, with $t_0 = 0$ (now) and $t_N = T$. (Note: from now on, we shall no longer use N to denote the number of assets.)

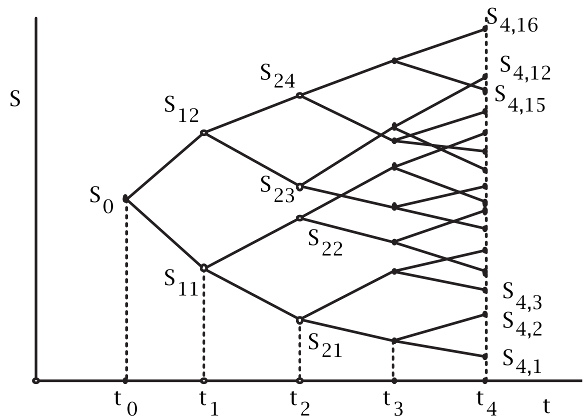

Figure 1.4: A non-recombining tree with $N = 4$ time levels.

Let us suppose that the interest rate is constant. If the bond has a final price $B_N = 1$ at $t = T$, then at $t = t_j$ it will have price $B_j = e^{-r(T-t_j)}$, independently of what the stock does.

1.2.1 Stock price model

At $t = t_0$ (now), the stock price has a single known value $S_0 = S_{00}$, that we obtain by calling our broker. Let us suppose that in the first time interval, between $t = t_0$ and $t = t_1$, the stock may move to either of two possible values, S_{11} or S_{12}. Here, the first subscript denotes time levels, the second one denotes possible price values.

We further suppose that starting from the value at t_1, there are exactly two new possible prices *for each* of the two price motions taken in the first period. There are thus *four* possible values the price can take at time t_2, and each one corresponds to a sequence of two successive motions. Thus after two periods have elapsed, one of four possible things will have happened; one of the following four trajectories will have occured:

$$
\begin{array}{ccccc}
S_0 & \to & S_{12} & \to & S_{24} \\
S_0 & \to & S_{12} & \to & S_{23} \\
S_0 & \to & S_{11} & \to & S_{22} \\
S_0 & \to & S_{11} & \to & S_{21}
\end{array}
$$

As time increases, each state splits into two states. At the end of N

periods, there are therefore 2^N possible states of the world. Each of these contains not only by a price $S_{N,j}$, for $j = 1, \ldots, 2^N$, but also by the entire history that led to that price. Such a model, with two branches at each node, is called a *binomial tree* (Figure 1.4).

For absence of arbitrage in this model, we need the prices on each "leaf" of the tree to satisfy the inequality constraint (1.2). Since the "children" of node (i, j) are $(i + 1, 2j - 1)$ and $(i + 1, 2j)$, we must apply it with the substitutions

$$S_0 \mapsto S_{i,j}, \qquad S_1 \mapsto S_{i+1,2j-1}, \qquad S_2 \mapsto S_{i+1,2j}, \qquad T \mapsto t_{i+1} - t_i$$

If this constraint were not satisfied at node (i, j), then we could construct an arbitrage strategy as follows: Do nothing until time t_i. If the stock price has moved to $S_{i,j}$, then carry out the arbitrage strategy at that time. That strategy generates a guaranteed profit by time t_{i+1}. It is not guaranteed that the price will reach $S_{i,j}$, but that event is possible, and hence this is an arbitrage strategy at the initial time.

1.2.2 Pricing a derivative

Here again, it gets interesting when we add an additional security. Suppose we add an option V. Suppose that the value of V is explicitly known at time t_N in terms of the price of S at that time. For example, V is a call option for which t_N is exactly the "exercise date," the time at which it can be exercised. Thus we know the values $V_{N,j}$ for each j.

Now we claim that we can use the pricing formula (1.3) to work backwards on the tree, all the way back to the current time t_0 and current stock price S_0. We first use the formula to determine the values $V_{N-1,*}$ at the next-to-last level in terms of the final values $V_{N,*}$. Then we determine the values $V_{N-2,*}$ in terms of the $V_{N-1,*}$. We repeat until we are left with the single number V_0, which is the price the option "ought" to be trading at in the market right now. By introducing multiple intermediate times at which trading is possible, we have been able to let the stock price take more than two values at the final time.

The reason this should be satisfied is as described above. If at any node on the tree this algebraic relation were *not* satisfied, then by waiting to see if the stock price actually moved to that node, we would have a finite chance of making a profit, with no chance of loss. And, provided V is fully known at $t = T$, the algebraic relations are enough to fix V_0.

1.2.3 Dynamic hedging

Along with the derivative value V, this algebra also gives us various auxiliary quantities at each node. Of these, the risk-neutral weight q has no direct meaning; we emphasize that it is not the probability of anything really happening in the world.

But the hedge ratios $\Delta_{i,j}$ are the key to the whole pricing strategy. The reason that the price of the option is uniquely defined at $t = 0$ is that there exists a *dynamic hedging strategy* that replicates its value at the final time. Thus the value of the option must be exactly equal to the cost of implementing the hedging strategy, or else there would be arbitrage opportunities.

In words, the strategy is the following. Suppose you sell an option to someone at time $t = 0$ for price V_0. At the same time, you go out and purchase Δ_0 shares of the stock; you deposit or borrow any left-over cash into or from an interest-bearing account.

At time t_1, the stock price will have moved either to $S_{1,1}$ or to $S_{1,2}$. Depending on which it has done, you adjust your stock position to $\Delta_{1,1}$ or to $\Delta_{1,2}$. This may yield some cash (if the new Δ is smaller than the old one) or require an input of cash. In either case you give or take to the interest-bearing account. In no case do you ever bring in cash from outside.

As time evolves and the stock price moves up and down, you continue this juggling act, continually rebalancing your position. At the final time, you will have exactly the right amount of stock and cash to cover your obligation to the person who bought the option, and you will be back to zero. (In practice, you add a markup to the initial option price to make yourself a sure profit.)

This strategy was invented in 1973 by Black, Scholes, and Merton; it is what they got the Nobel prize for (in the continuous-time limit).

Note that you have to continually adjust your stock holdings as the stock price changes. We can therefore add another to our list of key assumptions that go into the Black-Scholes pricing theory:

- *You can rebalance your portfolio as frequently as prices move.* Since prices generally move extremely rapidly, it is impractical to carry out this strategy exactly as described. Imperfect hedging leads to nonvanishing risk associated with writing an option.

Despite the immense simplifications that go into it, the Black-Scholes strategy has a tremendous importance. It brings the problem of deter-

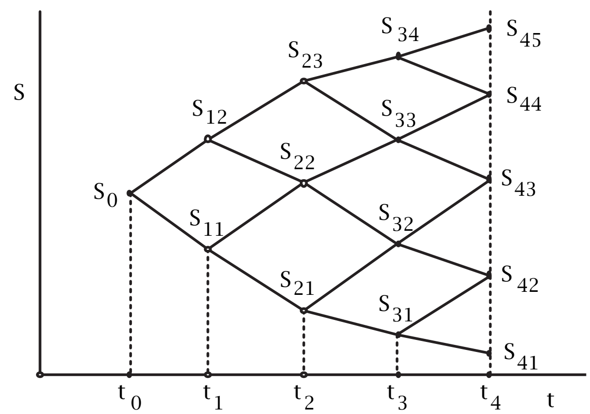

Figure 1.5: A recombining tree with $N = 4$ time levels.

mining an option price down from the realm of speculation, into the framework of something you can study rationally if imperfectly.

1.2.4 Recombining trees

One final wrinkle needs to be added. As you may already have observed and been wondering about, a tree as described above and as pictured in Figure 1.4 is extremely impractical. With N time levels, it has 2^N nodes at the final time. If N is large enough to achieve a reasonable degree of time resolution, this is an astronomical number. Furthermore, these values overlap in a horrendous way.

For practical purposes (both numerical and analytical computation) it is more reasonable to require the tree to be *recombining* as shown in Figure 1.5. In such a tree, we require an up motion followed by a down motion to give the *same price* as a down motion followed by an up motion.

A recombining tree has only $\mathcal{O}(N^2)$ elements for N time levels, so it is possible to achieve reasonable refinement in both time and stock price. By constructing this tree, we are assuming that the stock price will jump up and down on a grid whose structure we know in advance.

Of course, it remains to be seen what any of this really *means.* We can assume whatever we like about the future motion of a financial asset, but the world and the marketplace are under no obligation to conform to our model. We need a reasonable characterization of exactly what features of the tree determine the final price that we compute (Chapter 3).