# ONE

## NO ARBITRAGE:

## THE FUNDAMENTAL THEOREM OF FINANCE

THIRTY YEARS AGO marked the publication of what has come to be known as the Fundamental Theorem of Finance and the discovery of risk-neutral pricing.[1] The earlier option pricing results of Black and Scholes (1973) and Merton (1973) were the catalyst for much of this work, and certainly one of the central themes of this research has been an effort to understand option pricing. History alone, then, makes the topic of these monographs—the neoclassical theory of finance—particularly appropriate. But, history aside, the basic theorem and its attendant results have unified our understanding of asset pricing and the theory of derivatives, and have generated an enormous literature that has had a significant impact on the world of financial practice.

Finance is about the valuation of cash flows that extend over time and are usually uncertain. The basic intuition that underlies valuation is the absence of arbitrage. An arbitrage opportunity is an investment strategy that guarantees a positive payoff in some contingency with no possibility of a negative payoff and with no initial net investment. In simple terms, an arbitrage opportunity is a money pump, and the canonical example is the opportunity to borrow at one rate and lend at a higher rate. Clearly individuals would want to take advantage of such an opportunity and would do so at unlimited scale. Equally clearly, such a disparity between the two rates cannot persist: by borrowing at the low rate and lending at the high rate, arbitragers will drive the rates together. The Fundamental Theorem derives the implications of the absence of such arbitrage opportunities.

It's been said that you can teach a parrot to be an economist if it can learn to say "supply and demand." Supply, demand, and equilibrium are the catchwords of economics, but finance, or, if one is being fancy, financial economics, has its own distinct vocabulary. Unlike labor economics, for example, which

[1] The central tenet of the Fundamental Theorem, that the absence of arbitrage is equivalent to the existence of a positive linear pricing operator (positive state space prices), first appeared in Ross (1973), where it was derived in a finite state space setting, and the first statement of risk neutral pricing appeared in Cox and Ross (1976a, 1976b). The Fundamental Theorem was extended to arbitrary spaces in Ross (1978) and in Harrison and Kreps (1979), who described risk-neutral pricing as a martingale expectation. Dybvig and Ross (1987) coined the terms "Fundamental Theorem" to describe these basic results and "Representation Theorem" to describe the principal equivalent forms for the pricing operator.

specializes the methodology and econometrics of supply, demand, and economic theory to problems in labor markets, neoclassical finance is qualitatively different and methodologically distinct. With its emphasis on the absence of arbitrage, neoclassical finance takes a step back from the requirement that the markets be in full equilibrium. While, as a formal matter, the methodology of neoclassical finance can be fitted into the framework of supply and demand, depending on the issue, doing so can be awkward and may not be especially useful. In this chapter we will eschew supply and demand and develop the methodology of finance as the implication of the absence of arbitrage.

## No Arbitrage Theory: The Fundamental Theorem

The assumption of no arbitrage (NA) is compelling because it appeals to the most basic beliefs about human behavior, namely that there is someone who prefers having more wealth to having less. Since, save for some anthropologically interesting societies, a preference for wealth appears to be a ubiquitous human characteristic, it is certainly a minimalist requirement. NA is also a necessary condition for an equilibrium in the financial markets. If there is an arbitrage opportunity, then demand and supply for the assets involved would be infinite, which is inconsistent with equilibrium. The study of the implications of NA is the meat and potatoes of modern finance.

The early observations of the implications of NA were more specific than the general theory we will describe. The law of one price (LOP) is the most important of the special cases of NA, and it is the basis of the parity theory of forward exchange. The LOP holds that two assets with identical payoffs must sell for the same price. We can illustrate the LOP with a traditional example drawn from the theory of international finance. If $s$ denotes the current spot price of the Euro in terms of dollars, and $f$ denotes the currently quoted forward price of Euros one year in the future, then the LOP implies that there is a lockstep relation between these rates and the domestic interest rates in Europe and in the United States.

Consider individuals who enter into the following series of transactions. First, they loan \$1 out for one year at the domestic interest rate of $r$, resulting in a payment to them one year from now of $(1 + r)$. Simultaneously, they can enter into a forward contract guaranteeing that they will deliver Euros in one year. With $f$ as the current one-year forward price of Euros, they can guarantee the delivery of

$$(1 + r)f$$

Euros in one year's time. Since this is the amount they will have in Euros in one year, they can borrow against this amount in Europe: letting the Euro interest rate be $r_e$, the amount they will be able to borrow is

$$\frac{(1+r)f}{(1+r_e)}.$$

Lastly, since the current spot price of Euros is $s$ Euros per dollar, they can convert this amount into

$$\frac{(1+r)f}{(1+r_e)s}$$

dollars to be paid to them today.

This circle of lending domestically and borrowing abroad and using the forward and spot markets to exchange the currencies will be an arbitrage if the above amount differs from the \$1 with which the investor began. Hence, NA generally and the LOP in particular require that

$$(1+r)f = (1+r_e)s,$$

which is to say that having Euros a year from now by lending domestically and exchanging at the forward rate is equivalent to buying Euros in the current spot market and lending in the foreign bond market.

Not surprisingly, as a practical matter, the above parity equation holds nearly without exception in all of the foreign currency markets. In other words, at least for the outside observer, none of this kind of arbitrage is available. This lack of arbitrage is a consequence of the great liquidity and depth of these markets, which permit any perceived arbitrage opportunity to be exploited at arbitrary scale. It is, however, not unusual to come across apparent arbitrage opportunities of mispriced securities, typically when the securities themselves are only available in limited supply.[2]

While the LOP is a nice illustration of the power of assuming NA, it is somewhat misleading in that it does not fully capture the implications of removing arbitrage opportunities. Not all arbitrage possibilities involve two different positions with identical cash flows. Arbitrage also arises if it is possible to establish two equally costly positions, one of which has a greater set of cash flows in all circumstances than the other. To accommodate this possibility, we adopt the following framework and definitions. While the results we obtain are quite general and apply in an intertemporal setting, for ease of exposition we will focus on a one-period model in which decisions are made today, at date 0, and payoffs are

[2] I once was involved with a group that specialized in mortgage arbitrage, buying and selling the obscure and arcane pieces of mortgage paper stripped and created from government pass-through mortgages (pools of individual home mortgages). I recall one such piece—a special type of "IO"—which, after extensive analysis, we found would offer a three-year certain return of 37 percent per year. That was the good news. The bad news was that such investments are not scalable, and, in this case, we could buy only \$600,000 worth of it, which, given the high-priced talent we had employed, barely covered the cost of the analysis itself. The market found an equilibrium for these very good deals, where the cost of analyzing and accessing them, including the rents earned on the human capital employed, was just about offset by their apparent arbitrage returns.

received in the future at date 1. By assumption, nothing happens in between these two dates, and all decisions are undertaken at the initial date, 0.

To capture uncertainty, we will assume that there is a state space, $\Omega$, and to keep the mathematics at a minimum, we will assume that there are only a finite number of possible states of nature:

$$\Omega = \{\theta_1, \ldots, \theta_m\}.$$

The state space, $\Omega$, lists the mutually exclusive states of the world that can occur, $m$. In other words, at time 1 the uncertainty is resolved and the world is in one and only one of the $m$ states of nature in $\Omega$.

We will also assume that there are a finite number, $n$, of traded assets with a current price vector:

$$p = (p_1, \ldots, p_n).$$

Lastly, we will let

$$(\eta_1, \ldots, \eta_n)$$

denote an arbitrage portfolio formed by taking combinations of the $n$ assets. Each element of $\eta$, say $\eta_i$, denotes the investment in asset $i$. Any such portfolio will have a cost,

$$p\eta = \sum_i p_i \eta_i,$$

and we will refer to such a combination, $\eta$, as an arbitrage portfolio if it has no positive cost,

$$p\eta \leq 0.$$

We represent the Arrow-Debreu tableau of possible security payoffs by

$$G = [g_{ij}] = [\textit{payoff of security } j \textit{ if state } \theta_{ij} \textit{ occurs}].$$

The rows of G are states of nature and the columns are securities. Each row of the matrix, G, lists the payoffs of the $n$ securities in that particular state of nature, and each column lists the payoffs of that particular security in the different states of nature.

With the previous notation, we can define an arbitrage opportunity.

**Definition:** An *arbitrage* opportunity is an arbitrage portfolio with no negative payoffs and with a positive payoff in some state of nature. Formally, an arbitrage opportunity is a portfolio, $\eta$, such that

$$p\eta \leq 0$$

and

$$G\eta > 0,$$

where at least one inequality for one component or the budget constraint is strict.[3]

We can simplify this notation and our definition of NA by defining the stacked matrix:

$$A = \begin{bmatrix} -p \\ G \end{bmatrix}.$$

**Definition:** An arbitrage is a portfolio, $\eta$, such that

$$A\eta > 0.$$

Formally, then, the definition of no arbitrage is the following.

**Definition:** The principle of *no arbitrage (NA)*:

$$NA \Leftrightarrow \{\eta \mid A\eta > 0\} = \varnothing,$$

that is, there are no arbitrage portfolios.

The preceding mathematics captures our most basic intuitions about the absence of arbitrage possibilities in financial markets. Put simply, it says that there is no portfolio, that is, no way of buying and selling the traded assets in the market so as to make money for sure without spending some today. Any portfolio of the assets that has a positive return no matter what the state of nature for the world in the future, must cost something today. With this definition we can state and prove the Fundamental Theorem of Finance.

**The Fundamental Theorem of Finance:** The following three statements are equivalent:

1. No Arbitrage (NA).
2. The existence of a positive linear pricing rule that prices all assets.
3. The existence of a (finite) optimal demand for some agent who prefers more to less.

**Proof:** The reader is referred to Dybvig and Ross (1987) for a complete proof and for related references. For our purposes it is sufficient to outline the argument. A linear pricing rule, $q$, is a linear operator that prices an asset when applied to that asset's payoffs. In this finite dimensional setup, a linear pricing rule is simply a member of the dual space, $R^m$ and the requirement that

[3] We will use "≥" to denote that each component is greater than or equal, ">" to denote that ≥ holds and at least one component is greater, and "≫" to denote that each component is greater.

it is positive is just the requirement that $q \gg 0$. The statement that $q$ prices the assets means simply that $q$ satisfies the system of equations:

$$p = qG.$$

It is easy to show that a positive linear pricing operator precludes arbitrage. To see this let $\eta$ be an arbitrage opportunity. Clearly,

$$0 \geq p\eta = (qG)\eta = q(G\eta).$$

But, since $q \gg 0$, this can only be true if

$$G\eta < 0,$$

which is inconsistent with arbitrage.

The converse result, that NA implies the existence of a positive linear pricing operator, $q$, is more sophisticated. Outlining the argument, we begin by observing that NA is equivalent to the statement that the set of net trades,

$$S = \{x \mid \exists \eta, x = A\eta\},$$

does not intersect the positive orthant since any such common point would be an arbitrage. Since $S$ is a convex set, this allows us to apply a separating hyperplane theorem to find a $y$ that separates $S$ from the positive orthant, $R^+$. Since

$$yR^+ > 0,$$

we have $y \gg 0$. Similarly, for all $\eta$, we have

$$yS \leq 0 \Rightarrow yA\eta \leq 0 \Rightarrow yA\eta = 0$$

(since if the inequality is strict, $-\eta$ will violate it, i.e., $S$ is a subspace.) Defining

$$q = (q_1, \ldots, q_m) = \frac{1}{y_1}(y_2, \ldots, y_{m+1})$$

implies that

$$p = qG,$$

hence $q$ is the desired positive linear pricing operator that prices the marketed assets.

Relating NA to the individual maximization problem is a bit more straightforward and constructive. Since any agent solving an optimization problem would want to take advantage of an arbitrage opportunity and would want to do so at arbitrary scale, the existence of an arbitrage is incompatible with a finite demand. Conversely, given NA, we can take the positive linear pricing rule, $q$, and use it to define the marginal utility for a von Neumann-Morgenstern expected utility maximizer and, thereby, construct a concave monotone utility function that achieves a finite maximum.

• • • • •

Much of modern finance takes place over time, a fact that is usually modeled with stochastic diffusion processes or with discrete processes such as the binomial. While our simplified statement of the Fundamental Theorem is inadequate for that setting, and while the theorem can be extended to infinite dimensional spaces, some problems do arise. However critical these difficulties are from a mathematical perspective, as a matter of economics their importance is not yet well understood. The nut of the difficulty appears to be that the operator that prices assets may not have a representation as a vector of prices for wealth in different states of nature, and that makes the usual economic analysis of trade-offs problematic.[4]

A *complete market* is one in which for every state $\theta_i$ there is a combination of the traded assets that is equivalent to a pure contingent state claim, in other words, a security with a payoff of the unit vector: one unit if a particular state occurs, and nothing otherwise. In a complete market G is of full-row rank, and the equation

$$p = qG$$

has a unique solution,

$$q = pG^{-1}.$$

This determinancy is one reason why market completeness is an important property for a financial market, and we will later discuss it in more detail. By contrast, in an *incomplete market*, the positive pricing operator will be indeterminate and, in our setting with $m$ states and $n$ securities, if $m > n$, then the operator will be an element of a subspace of dimensionality $m - n$. This is illustrated in figure 1.1 for the $m = 3$ state, $n = 2$ security example. In figure 1.1, each of the two securities has been normalized so that $R^1$ and $R^2$ represent their respective gross payoffs in each of the three states per dollar of investment.

## The Representation Theorem

The positive linear operator that values assets in the Fundamental Theorem has several important alternative representations that permit useful restatements of the theorem itself.

[4] Mathematically speaking, separation theorems require that the set of net trades be "fat" in an appropriate sense, and in, say, the $L^2$ norm, the positive orthant lacks an interior. This prevents the application of basic separation theorems and requires some modifications to the definition of arbitrage and no arbitrage (see Ross [1978a], who extends the positive linear operator by finessing this problem, and Harrison and Kreps [1979], who find a way to resolve the problem).

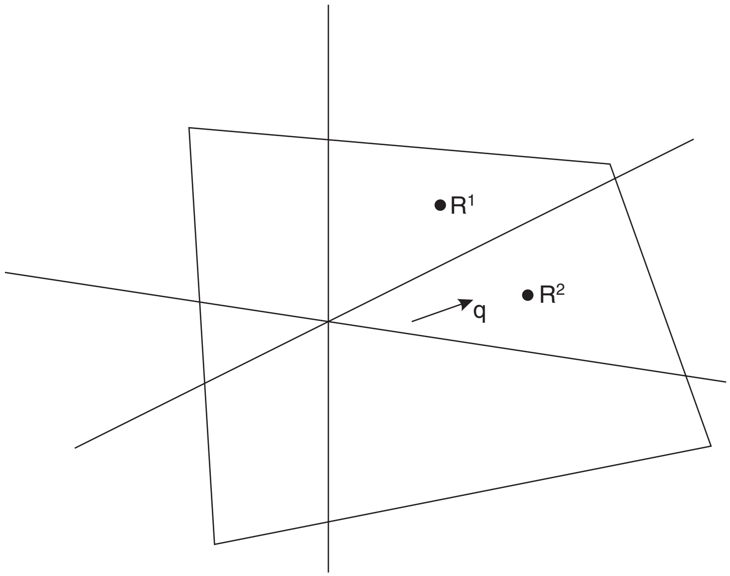


Figure 1.1

**Definition:** *Martingale* or *risk-neutral probabilities* are a set of probabilities, $\pi^*$, such that the value of any asset with payoffs of

$$z = (z_1, \ldots, z_m)$$

is given by the discounted expectation of its payoffs using the associated riskless rate to discount the expectation,

$$V(z) = \frac{1}{1+r}E^*[z] = \frac{1}{1+r}\sum \pi_i^* z_i.$$

**Definition:** A *state price density*, also known as a pricing *kernel*, is a vector,

$$\phi = (\phi_1, \ldots, \phi_m),$$

such that the value of any asset is given by

$$V(z) = E[\phi z] = \sum \pi_i \phi_i z_i.$$

These definitions provide alternative forms for the Fundamental Theorem.

**Representation Theorem:** The following statements are equivalent:

1. There exists a positive linear pricing rule.
2. The martingale property: the existence of positive risk-neutral probabilities and an associated riskless rate.
3. There exists a positive state price density or pricing kernel.

**Proof:**

**(1) ⇔ (2)**

From the Fundamental Theorem we know that there exists a positive pricing vector, $q \gg 0$, such that

$$V(z) = qz.$$

Consider, first, the sum of the $q_i$

$$V(e) = qe = \sum q_i.$$

We have written this as $V(e)$, where $e$ is the vector of 1 since it is the value of receiving 1 for sure, that is, of getting 1 in each state. We can define a rate of return, $r$, as the return from investing in this security, thus,

$$r = \frac{1}{V(e)} - 1$$

or

$$V(e) = \frac{1}{1+r} = qe = \sum q_i.$$

Notice that $r$ is uniquely defined if a riskless asset is traded explicitly or implicitly, that is, if $e$ is in the span of the marketed assets.

Next we define the risk-neutral probabilities as

$$\pi_i^* = \frac{q_i}{\sum q_i} > 0.$$

Notice that, just as probabilities should, the $\pi_i^*$ sum to 1.

Hence, we have

$$\begin{aligned} V(z) = \sum q_i z_i &= \left(\sum q_i\right) \sum \left( \frac{q_i}{\sum q_i} \right) z_i \\ &= \left(\frac{1}{1+r}\right) \sum \pi_i^* z_i = \left(\frac{1}{1+r}\right) E^*[z], \end{aligned}$$

where the symbol $E^*$ denotes the expectation taken with respect to the risk-neutral probabilities (or "measure").

Conversely, if we have a set of positive risk-neutral probabilities and an associated riskless rate, $r$, it is clearly a positive linear operator on $R_m$, and we can simply define the state space price vector as

$$q_i = \left(\frac{1}{1+r}\right)\pi_i^* > 0.$$

**(1) ⇔ (3)**

Defining the positive state density as

$$\phi_i = \frac{q_i}{\pi_i} > 0,$$

we have

$$\begin{aligned} V(z) &= \sum q_i z_i = \sum \pi_i \left(\frac{q_i}{\pi_i}\right) z_i \\ &= \sum \pi_i \phi_i z_i = E[\phi z], \end{aligned}$$

and the converse is immediate.

• • • • •

Much of the research in modern finance has turned on extending these results into mathematically deeper terrain. The martingale, or risk-neutral, approach in particular has been the focus of much attention because of the ability to draw on the extensive probability literature. Interestingly, the risk-neutral approach was discovered independently of the Fundamental Theorem (see Cox and Ross 1976a, 1976b) in the context of option pricing, and only later was the relation between the two fully understood.

Given the tradition of Princeton in the physical sciences and given the role of finance as, perhaps, the most scientific of the social sciences, it is only appropriate to point out the close relation between these results and some fundamental results in physics. Without much imagination, the Representation Theorem can be formally stated in a general intertemporal setting in which a security with a random payoff, $z_T$, across time, $T$, has a current value, $p$, given by

$$p = E^*[e^{-\int_0^T r_s ds} z_T],$$

where $r_s$ denotes the instantaneous risk-free interest rate and the asterisk denotes the expectation under the martingale measure. This is the usual modern statement of risk-neutral pricing. In this computation, the short interest rate can be stochastic, in which case all likely paths must be considered. More generally, too, the payoff itself could be path dependent as well. In practice, this expectation can be computed numerically by a summation over Monte Carlo simulations generated from the martingale measure. Such an integral, generated over paths, is equivalent to the Feynman-Kac integral of quantum mechanics.

## Option Pricing

### *Black-Scholes-Merton*

In the previously shown form, risk-neutral pricing (Cox and Ross 1976a, 1976b) provides the solution to the most famous and important examples of the implications of NA: the classic Black-Scholes-Merton (1973) option-pricing formula and the binomial formula of Cox, Ross, and Rubinstein (1979).[5] Without going into great detail, we will outline the argument.

The time $T$ payoff for a call option on a stock with a terminal price of $S_T$ at the option's maturity date, $T$, is:

$$C(S,\ T) = \max\{S_T - K, 0\},$$

where $K$ is referred to as the strike or exercise price.[6] To compute the current value of the option we take the expectation of this payoff under the risk-neutral measure, that is, the martingale expectation. This is exactly the risk-neutral valuation that is obtained by assuming that the expected return on the stock is the risk-free interest rate, $r$, and taking the ordinary expectation.

The Black-Scholes model assumes that the stock process follows a lognormal distribution. In the formalism of Ito's stochastic calculus, this is written as

$$dS = \mu\, dt + \sigma\, dz,$$

where $\mu$ is the local drift or expected return, $\sigma$ is the local speed or standard deviation, and the symbol $dz$ is interpreted as a local Brownian process over the interval $[t,\ t + dt]$ with mean zero and variance $dt$. Assuming that the call option has a current value of $C(S,\ t)$, where we have written its arguments to emphasize the dependence of value on the current stock price, $S$, and the current time, $t$, and applying the integral valuation equation, we obtain

$$C(S, 0) = E^*[e^{-r(T-t)} \max\{S_T - K, 0\}].$$

The approach of Black and Scholes was less direct. Using the local equation, they observed that the return on the call, that is, the capital gain on the call, would be perfectly correlated locally with the stock since its value depends on

[5] See Black and Scholes (1973) and Merton (1973) for the original analysis with a diffusion process, and Cox, Ross, and Rubinstein (1979) for the binomial pricing model.

[6] We are assuming that the call is on a stock that pays no dividends and that it is a European call, which is to say that it can be exercised only at maturity and not before then.

the stock price. They used this result together with a local version of the Capital Asset Pricing Model (discussed later) to derive a differential equation that the call value would follow, the Black-Scholes differential equation:

$$\tfrac{1}{2}\sigma^2 S^2 C_{SS} + rSC_S + rC = -C_t.$$

Notice the remarkable fact that the assumed expected return on the stock, $\mu$—so difficult to measure and surely a point of contention among experts—plays no role in determining the value of the call.

Merton pointed out that since the returns on the call were perfectly correlated with those on the underlying stock, a portfolio of the risk-free asset and the stock could be constructed at each point in time that would have identical returns to those on the call. To prevent arbitrage, then, a one-dollar investment in this replicating portfolio would be equivalent to a one-dollar investment in the call. The resulting algebra produces the same differential equation, and, as a consequence, there is no need to employ an asset-pricing model. Black and Scholes and Merton appealed to the literature on differential equations to find the solution to the boundary value problem with the terminal value equal to the payoff on the call at time $T$.

Cox and Ross observed that once it was known that arbitrage alone would determine the value of the call option, analysts were handing off the problem to the mathematicians prematurely. Since arbitrage determines the value, they argued that the value would be determined by what they called risk-neutral valuation, that is, by the application of the risk-neutral formula in the Representation Theorem. In fact, the Black-Scholes differential equation for the call value is the backward equation for the log-normal stochastic process applied to the risk-neutral expected discounted value integral, and, conversely, the discounted integral is the solution to the differential equation. The assumed expected return on the stock, $\mu$, is irrelevant for valuation since in a risk-neutral world all assets have the riskless rate as their expected return.

It follows that the value is simply the discounted expected value of the terminal payoff under the assumption that the drift on the stock is the risk-free rate. Applying this analysis yields the famous Black-Scholes formula for the value of a call option:

$$C(S, 0) = SN(d_1) - e^{-rT} KN(d_2),$$

where $N(\cdot)$ denotes the standard cumulative normal distribution function and where

$$d_1 = \frac{\ln(S/K) + (r + \tfrac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$$

and

$$d_2 = d_1 - \sigma T.$$

### *The Binomial Model*

The crucial features of the Black-Scholes-Merton analysis were not immediately apparent until the development of the binomial model by Cox, Ross, and Rubinstein. Was it the lognormal distribution, the continuity that allowed continuous trading or some other feature of the model, and would the analysis fall apart without these features? The binomial model resolved these questions, and because of its flexibility it has now become the preferred structure for valuing complex derivatives in the practical world of finance.

The binomial analysis is illustrated in figure 1.2. The value of the option at time $t$ is given by $C(S, t)$ where we explicitly recognize its dependence on $S$, the stock price at time $t$. There are two states of nature at each time, $t$, state $a$ and state $b$, representing the two possible futures for the stock price, $aS$ or $bS$ where $a > 1 + r > b$. The figure displays the gross returns on the three assets, $1 + r$ for the bond, $a$ or $b$ for the stock, and the formula $C(aS, t + 1)/C(S, t)$ for the option if state $a$ occurs and $C(bS, t + 1)/C(S, t)$ if $b$ occurs. With two states and three assets, one of the assets may be selected to be redundant, and this is represented by the line through the gross returns of the assets. The line is the combination of returns across the states, available by combining a unit investment in the three assets. To prevent arbitrage these three points must

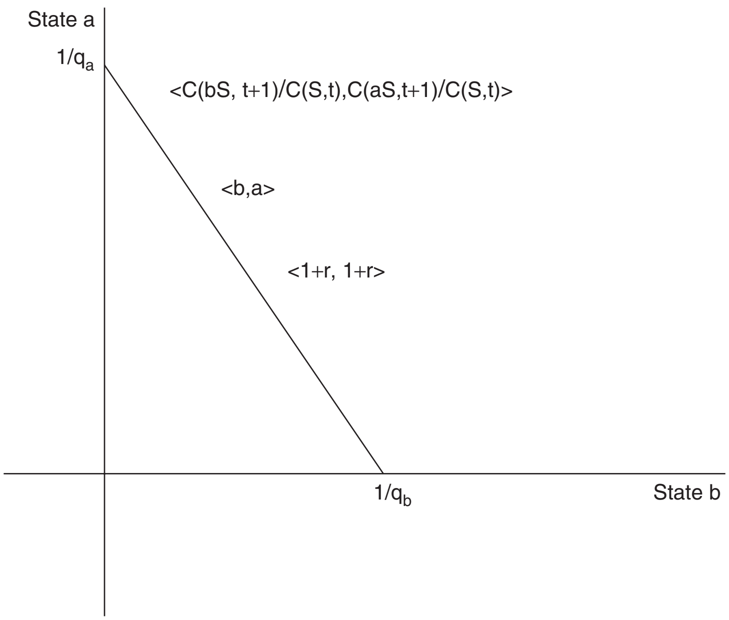


Figure 1.2

all lie on a line or else a portfolio of two of them would dominate the third. The equation of this line provides the fundamental difference equation, which can be solved subject to the boundary condition of the payoff at maturity, $T$, to give the current value of the option:

$$(1+r)C(S,\tau)=\frac{1+r-b}{a-b}C(aS,\tau-1)+\frac{a-1+r}{a-b}C(bS,\tau-1),$$

where $\tau$ is the number of periods left to maturity, $T-t$. The coefficients of the next period values in the equation are called the risk neutral probabilities for the binomial process, that is, they are the probabilities for the process under the assumption that the expected return is the risk-free rate. Notice that parallel with the absence of a role for the drift in the diffusion model, because the market is locally complete with two states and three assets, one of the assets is redundant and, given the values of the other two, the probabilities for the binomial jumps are not needed to value the third security. If we let the time difference between nodes grow small and increase in number so as to maintain a constant volatility for the stock return, this difference equation converges to the Black-Scholes differential equation.

Alternatively, we can derive the difference equation directly from the Fundamental Theorem by invoking NA directly to price a derivative security such as a call option. Since the stock and the option are perfectly correlated assets locally, we can form a portfolio of the two of them that is riskless, for example, we can short the stock and go long the option in just the correct proportions so as to eliminate all risk in the portfolio. Equivalently, we could construct a replicating portfolio composed of the risk-free rate and the stock whose return is locally identical to the return on the call option. NA in the form of the LOP requires that the portfolio, as a riskless asset, have the unique risk-free rate of return, $r$. The resulting equation is the discrete equation of the binomial model and the famous partial differential equation of Black and Scholes. In either case, the solution is the pricing equation derived from risk-neutral pricing.[7]

What, then, are the crucial features that allow for arbitrage to determine the price of the derivative security or option as a function of traded securities? Clearly, local spanning is sufficient (and except in unusual cases it is necessary) in the sense that the return of the derivative security must be a (locally) linear combination of the returns on the stock and any other marketed securities that are required to span its return. For example, if the volatility of the

[7] While the above is the most straightforward derivation, it is not the most powerful. In fact, there is no need to assume that the option inherits the stochastic properties of the underlying asset since it is possible to replicate the payoff of the option with a portfolio of the stock and the bond. Through NA this would force the value of the option to be the original cost of such a portfolio. See Ingersoll (1987) for this argument.

stock is itself random, then another instrument such as, say, the movement on a traded volatility index or on a market index that is perfectly correlated with changes in the volatility, would have to be used; and the value of the option would depend not only on the stock on which it was directly written, but also on the value of these other securities. As another example, we could have an option whose payoff depended not simply on the return of a single stock but rather on two or more assets, $S$ and $P$ with a payoff at time $T$ of

$$S^2 + P.$$

In response to such complications, in discrete time the binomial might give way to a trinomial or an even more complex process (see Cox and Ross 1976a, 1976b).

## Some Further Applications of the Representation Theorem

### *Asset Pricing*

Historically and currently the pricing of assets has been a focus of neoclassical finance and, not surprisingly, a number of asset-pricing theories have been developed. While these differ in their form, they share some common intuitions. The theory of asset pricing began hundreds of years ago with a simple formulation drawn from the intuitions of gambling. The common sense view was that an asset would have a "fair" price if its expected return was simply the riskless rate,

$$E_i \equiv E[x_i] = r,$$

where $x_i$ denotes the expected rate of return on asset $i$.[8] If this were so, then, on average, the asset would neither excessively reward nor penalize the holder.

With the work of Hicks and Samuelson and others, it was recognized that this was not an adequate representation. After all, individuals are generally assumed to be averse to risk and, therefore, assets whose returns are risky should command a lower price than riskless assets. This was captured by positing a return premium, $\pi_i$, for risky assets,

$$E_i = r + \pi_i,$$

where $\pi_i > 0$ to capture the need for additional return to compensate for risk.

Both theoretical and empirical works have focused on the determinants of the risk premium. Presumably $\pi$ should be determined by two effects. On the one hand, the riskier the asset the greater the premium should be for bearing that risk and, on the other hand, the more averse investors are to risk, the greater the premium they would demand for bearing that risk. If we let risk

[8] This is usually credited to Bernoulli.

aversion be captured by a parameter, $R$, which could denote some average of the von Neumann-Morgenstern risk-aversion parameter for individual agents, and if we represent uncertainty by the simple variance of the asset return, $\sigma^2$, then we would have

$$\pi_i \propto R\sigma_i^2$$

with, perhaps, other factors such as asset supply and the like entering in.

It took decades from this initial set of concepts to the development of mean variance portfolio theory and the intuitive leap of the Capital Asset Pricing Model (CAPM) to recognize that this way of thinking about the risk premium was not quite correct. The contribution of modern neoclassical theory to this question comes with the recognition that since an asset fits into a portfolio, what matters for determining its risk premium is how it relates to the other assets in that portfolio, and not simply its overall volatility. We can illustrate this with a look at the three main asset-pricing theories.

### *Arbitrage Pricing Theory (APT)*

The Arbitrage Pricing Theory (APT) is the most direct descendant of the Representation Theorem (see Ross 1973, 1976a). It begins with the simple observation that if one wants to say more about asset returns than can be obtained from arbitrage alone, one must make some assumptions about the distribution of returns. Suppose, then, we posit that returns follow an exact linear model,

$$x_i = E_i + \beta_i f,$$

where $f$ is a common factor (or vector of factors) that influences the returns on all assets and $\beta_i$ is the beta, or loading, for asset $i$ on that factor. The vector $f$ captures the innovations in the state variables that influence the returns of all the assets and, mathematically, in a static setting this is equivalent to a rank restriction on the matrix of returns.

Since all the assets are perfectly correlated, in this exact one-factor world we would expect the force of NA to dictate a pricing result. In fact, the logic is identical to that which we employed to derive the binomial pricing model for options. Since a unit investment in any asset gives a payoff equal to the gross return on that asset, the value of any gross return must be 1,

$$\begin{aligned} 1 &= \frac{1}{1+r} E^*[1 + x_i] \\ &= \frac{1}{1+r}(1 + E^*[E_i + \beta_i f]) \\ &= \frac{1 + E_i + \beta_i E^*[f]}{1+r}. \end{aligned}$$

Rearranging the equation, we obtain the familiar statement of the risk premium, the Security Market Line (SML):

$$E_i - r = -E^*[f]\beta_i = \pi_f \beta_i,$$

where

$$\pi_f \equiv -E^*[f].$$

Observing from the SML that $\pi_f$ is the risk premium on an asset with a beta of 1, we can rewrite the SML in the traditional form as

$$E_i - r = (E_f - r)\beta_i,$$

where $E_f$ denotes the expected return on any portfolio with a beta of 1 on the factor. In other words, the risk premium depends on the beta of the asset that captures how the asset relates—or correlates—to the other assets through its relation to the common factor(s) and the premium, $\pi_f$, on those factor(s).

The exact specification of the foregoing statement is too stark for empirical practice because outside of option pricing and the world of derivative securities, assets such as stocks are not perfectly correlated and are subject to a host of idiosyncratic influences on their returns. We can modify the return generating mechanism to reflect this as

$$x_i = E_i + \beta_i f + \varepsilon_i,$$

where $\varepsilon_i$ is assumed to capture the idiosyncratic, for example, company- or industry-specific forces on the return of asset *i*. If the idiosyncratic terms are sufficiently independent of each other, though, by the law of large numbers asymptotically as the number of assets in a well-diversified portfolio is increased, the portfolio return will approach the original exact specification. Such arguments lead to the SML holding in an approximate sense.[9]

While the intuition of the APT is clear, namely that factors exogenous to the market move returns and that pricing depends on them, and, further, that we can capture these by using endogenous variables such as those created by forming portfolios with unit loadings on the factors, exactly what these factors might be is unspecified by the theory. Next we turn to the traditional Capital Asset Pricing Model (CAPM) and its cousin, the Consumption Beta Model (CBM), to get more explicit statements about the SML.

[9] These conditions have led to a wide literature and some debate as to exactly how this return is accomplished and what is to be meant by an approximate pricing rule. See, for example, Shanken (1982) and Dybvig and Ross (1985).

### *The Capital Asset Pricing Model (CAPM) and the Consumption Beta Model (CBM)*

Historically, the CAPM (Sharpe 1964; Lintner 1965) preceded the APT, but it is a bit less obvious what its relation is to NA. By assuming that returns are normally distributed (or locally normal in a continuous time-diffusion model), or by assuming that individuals all have quadratic utility functions, it is possible to develop a beautiful theory of asset pricing. In this theory two additional concepts emerge. First, since the only focus is on the mean and the variance of asset returns, individuals will choose portfolios that are mean-variance efficient, that is, portfolios that lie on the frontier in mean-variance space. This is equivalent to what is called two-fund separation (see Ross [1978b], who observes that all efficient portfolios lie on a line in the *n*-dimensional portfolio space). Second, a consequence of mean-variance efficiency is that the market portfolio, $m$, that is, the portfolio in which all assets are held in proportion to their values, will itself be mean-variance efficient. Indeed, the mean-variance efficiency of the market portfolio is equivalent to the CAPM.[10]

Not surprisingly, in any such model where valuation is by a quadratic, the pricing kernel will be in terms of marginal utilities or derivatives of a quadratic and will be linear in the market portfolio. It can be shown that the pricing kernel for the CAPM[11] has the form

$$\varphi = \frac{1}{1+r}[1-\lambda(m-E_m)].$$

Hence any asset must satisfy

$$\begin{aligned}1 &= E[\varphi(1+x_i)] \\ &= E\left[\left(\frac{1}{1+r}[1-\lambda(m-E_m)](1+x_i)\right)\right] \\ &= \frac{1+E_i}{1+r} - \frac{\lambda}{1+r}\mathrm{cov}(x_i, m).\end{aligned}$$

Setting $i = m$, we can solve for $\lambda$,

$$E_m - r = \lambda\sigma_m^2,$$

[10] See Stephen Ross (1977). Richard Roll (1977) pointed out the difficulties that the potential inability to observe the entire market portfolio raises for testing the CAPM.

[11] The astute reader will recognize that this operator is not positive, and, therefore, the CAPM admits of arbitrage outside of the domain of assets to which the model is deliberately restricted. See Dybvig and Ingersoll (1982) for a discussion of these issues.

which allows us to rearrange the pricing equation to the SML:

$$\begin{aligned} E_i - r &= \lambda \operatorname{cov}(x_i, m) \\ &= (E_m - r)\beta_i, \end{aligned}$$

where $\beta_i$ is the regression coefficient of asset $i$'s returns on the market, $m$.

The SML verifies a powerful concept. Individuals who hold a portfolio will value assets for the marginal contributions they will make to that portfolio. If an asset has a positive beta with respect to that portfolio, then adding it into the portfolio will increase the volatility of the portfolio's returns. To compensate a risk-averse investor for this increase in volatility, in equilibrium such an asset must have a positive excess expected return, and the SML verifies that this excess return will, in fact, be proportional to the beta.

This is all valid in a one-period world where terminal wealth is all consumed and end-of-period consumption is the same as terminal wealth. Once we look at a multiperiod world, the possibility of intermediate consumption separates the stock of wealth from the flow of consumption. Not surprisingly, though, with a focus only on consumption and with a restriction that preferences depend only on the sum of the utilities of the consumption flow, the SML again holds, with $\beta_i$ as the regression coefficient of asset $i$ on individual or, when aggregation is possible, on aggregate consumption, and with the $E_m$ interpreted as the expected return on a portfolio of assets that is perfectly correlated with consumption. This result is called the Consumption Beta Model (CBM) (see Merton 1971; Lucas 1978; and Breeden 1979).

More generally, in the next chapter we will exploit the fact that the Representation Theorem allows pricing to take the form of the security market line where the market portfolio is replaced by the pricing kernel, that is, excess expected return on assets above the risk-free return is proportional to their covariance with the pricing kernel. In a complete market where securities are traded contingent on all possible states and where agents have additive separable von Neumann-Morgenstern utility functions, individuals order their consumption across states inversely to the kernel, which implies that aggregate consumption is similarly ordered. This fact allows us to use aggregate consumption as a state variable for pricing and replaces covariance with the pricing kernel in the SML with covariance with aggregate consumption.

## Corporate Finance

Corporate finance is the study of the structure and the valuation of the ownership claims on assets through various forms of business enterprise. The famous Modigliani-Miller (MM) Theorem (Modigliani and Miller 1958) on the irrelevance of the corporate financial structure for the value of the firm emerges as a corollary of the NA results. The MM Theorem is to modern corporate finance what the famous Arrow Impossibility Theorem is to the Theory of Social Choice. Like a great boulder in our path, it is too big to move and all of current research can be interpreted as an effort to circumvent its harsh implications. The MM Theorem teaches us that in the pristine environment of perfect markets with no frictions and no asymmetries in information, corporate finance is irrelevant. As with the Impossibility Theorem, appealing axioms have seemingly appalling consequences.

Consider a firm with an array of claims against its value. Typically we model these claims as debt securities to which the firm owes a determined schedule of payoffs with bankruptcy occurring if the firm cannot pay these claims in full. The residual claimant is equity, which receives the difference between the value of the firm at time 1 and the payoffs to the debt holders provided that the difference is positive, and which receives nothing if the firm is in bankruptcy. Notice that the equity claim is simply a call on the value of the firm with an exercise price equal to the sum of the face value of the debt claims. The statement of the MM Theorem, though, is much more general than this particular case.

**The Modigliani-Miller (MM) Theorem:** The value of a firm is independent of its financial structure.

**Proof:** Suppose a firm has a collection of claims against its value, with payoffs of $\langle z^1, \ldots, z^k \rangle$ at time 1. Since we have included all of the claimants, we must have

$$z^1 + \cdots + z^k = V_1.$$

As a consequence, at time 0 we have

$$\begin{aligned} \textit{value of firm} &= L(z^1) + \cdots + L(z^k) \\ &= L(z^1 + \cdots + z^k) \\ &= L(V_1), \end{aligned}$$

which is independent of the particulars of the individual claims.

• • • • •

## Conclusion

There is, of course, much more to neoclassical finance than the material or even the topics presented in this chapter. We have omitted a detailed examination of individual portfolio theory and have hardly mentioned the fascinating intertemporal analysis of both individual optimal consumption and portfolio-choice problems as well as the attendant equilibria. We will rectify a small portion of this in the next chapter, but we will still not have time or space adequately to treat the intertemporal analysis. Having read that, though, a cynic might observe that almost all of contemporary intertemporal analysis is really just an extension of the static analysis of this chapter.

This coincidence occurs for two rather opposite reasons. On the one hand, most intertemporal models are simply sequences of one-period static models. With some rare exceptions, there are really no true intertemporal models in which consumption or production interacts over time in a fashion that is not additively separable into a sequence of static models. Secondly, a marvelous insight by Cox and Huang (1989) demonstrates that in a complete market the optimization and equilibrium problems may be treated as though the world were static—the integral view. If the market is complete, then any future pattern of consumption can be achieved by choosing a portfolio constructed from the array of state space-contingent securities. Such a portfolio will have a representation as a dynamic policy and it will produce the desired intertemporal consumption profile. In a sense, if the number of states of a static problem is augmented sufficiently, it can replicate the results of a complete intertemporal world. We will expand a bit on these issues in the next chapter.