# Chapter 2

# Arbitrage

Financial futures, and derivatives in general, are built on pre-existing *underlying* securities. By simultaneously conducting operations on several markets, specialized intermediaries intervene whenever an *opportunity for arbitrage* arises that ensures a profit in all contingencies. With the increasing sophistication of derivatives, the ancient art of arbitrage can become very complicated today. These interventions ensure some price consistency across markets. In particular, they induce relationships between the prices of securities and derivatives that lead to the procedures of *valuation by arbitrage* that are systematically used by financial institutions.

The goal of this chapter is to formalize and analyze the notion of arbitrage. The underlying assumptions and the limits of the arbitrage-based valuation procedures that are used by financial institutions are made explicit.

Uncertainty is described in terms of *states of nature* that determine future payoffs. An *opportunity for arbitrage* consists of transactions in which no money can be lost and some can be earned in certain states of nature. When there are no frictions, such opportunities should not last, which motivates the study of markets without arbitrage opportunities. The absence of arbitrage opportunities dictates some relationships between the prices of securities and their payoffs that are easily expressed in terms of *state prices*. These relationships also allow the *valuation* of some securities on the basis of the prices of other securities. This procedure however is only valid under certain conditions. In particular, a natural and key distinction is made between *complete* markets, for which the valuation procedure is always valid, and markets which are *incomplete*.

The first section studies a static framework. In the second section, the analysis is extended to the dynamic framework that underlies the most commonly used valuation methods, at the cost of strong assumptions on expectations. We emphasize that arbitrage relationships only allow to get the prices of some derivatives from others, but never determine the prices of the whole set of securities: such a determination typically requires a complete economic model, similar to those of Chapters 5 and 6.

# 1 Static arbitrage

#### 1.1 States of nature

Consider an economy at time 0, with a future that may be uncertain. In this section, the future is reduced to a single point in time, t = 1. Uncertainty is modeled by a set of *states of nature*, which represents all possibilities at time 1. A *state of nature* e provides a description of the economic environment: It includes all relevant information, such as agents' tastes, their resources, firms' profits, dividends paid by each asset, etc. We assume here that the number of states is finite, with  $\mathcal{E}$  denoting the set of states and E representing the number of states in  $\mathcal{E}$ .

The appropriate set of states depends on the problem under investigation and may be more or less complex. The only constraint is that the characteristics of the economy can be expressed as a deterministic function of the state. For example, to price a European call option, the state may be summarized by the price of the underlying security, S, (assumed to take a finite number of values), because the payoff accruing to the holder of this option is a function of that price,  $\max(S - K, 0)$ .

The set of states of nature is analogous to the fundamental space of probability theory. However, notice that throughout this chapter we never say that some states are more or less probable than others. No probability distribution is specified over the state space. Indeed, an opportunity for arbitrage arises when operations can be conducted that yield profits in some states of nature without generating losses in any other state. The existence, or absence, of any such opportunity depends exclusively on the set of possible states, and not on any probability distribution on these states.

#### 1.2 Securities

Consider a market, opened at time 0, on which are traded k = 1, ..., Ksecurities. A unit of security k is defined by a payoff (coupons, dividends, resale price at some future date) in dollars, which the owner receives in the various states of nature:  $a_k(e)$  represents the payoff contingent on the occurrence of state e, e in  $\mathcal{E}$ .

The matrix of the securities payoffs is denoted  $\tilde{a}$ , its elements are  $a_k(e)$ and its dimension is  $(K \times E)$ . The price today of security k is designated  $p_k$ , and the vector of the prices p is a column vector in  $\mathbb{R}^K$ .

The set of all states, the contingent payoffs, and the prices of the securities are the data characterizing the markets. They are summarized by  $(\mathcal{E}, \tilde{a}, p)$ .

A portfolio specifies the (positive) holdings, or long positions, as well as the debts (negative), or short positions in the various securities. It is represented by a column vector,  $z = (z_k)_{k=1,...,K}$ , the k-th element of which,  $z_k$ , indicates the number of units of security k in the portfolio when it is positive. If it is negative, the portfolio is *short* on security k which commits its owner to paying  $|z_k|a_k(e)$  at time t = 1 in the event that state e materializes.

The *value* of the portfolio is equal to:

$$\sum_{k=1}^{K} p_k z_k = p' z.$$

The portfolio yields in state e the payoff, also called revenue or income:

$$c_z(e) = \sum_{k=1}^K a_k(e) z_k.$$

The vector of contingent payoffs associated with z, a row vector in  $\mathbb{R}^{E}$ , is simply written

$$\tilde{c}_z = z'\tilde{a}.$$

We shall use this convention throughout the book: securities prices, portfolios are represented by column vectors, contingent payoffs by row vectors.

This representation encompasses financial instruments that obligate their owners to pay out in some states of nature, such as bets on elections. Formally, some of the payoffs  $a_k(e)$  may be negative.

Consider for example a forward market on a commodity, say wheat, opened at time 0 and maturing at time 1. Recall that at maturity the payoff of the forward contract is the difference between the price of wheat on the spot market, and the forward price, f, determined at time 0.

Let w(e) denote the spot price of wheat at time 1 where e is the state that materializes. If we abstract from guarantee deposits, no payments are made when the contract is signed. At maturity the buyer of a contract gets the (positive or negative) payoff w(e) - f, if e materializes. Thus, in terms of our conventions, a forward contract corresponds to a financial instrument k, the price of which today is nil,  $p_k = 0$ , and which yields the contingent payoff  $a_k(e) = w(e) - f$ , tomorrow. Typically, there are states in which the spot price exceeds the futures price, and vice versa.

#### Example 1

There are two states of nature and two securities. a risk-free security with rate r, and a stock whose price<sup>1</sup> S at t = 0 can move to (1 + h)S in one

<sup>&</sup>lt;sup>1</sup>This model will be used and developed to price an option on the stock. Thus we use the standard notation of option models: S denotes the price of the underlying security.

state and (1+b)S in the other. The model is written:

$$\mathcal{E} = \{h, b\}, \ p = \begin{bmatrix} 1\\ S \end{bmatrix} \ \tilde{a} = \begin{bmatrix} 1+r & 1+r\\ (1+h)S & (1+b)S \end{bmatrix}$$

The payoff of portfolio z is:

$$[z_1(1+r) + z_2(1+h)S, z_1(1+r) + z_2(1+b)S].$$

Figure 1 represents the case of a stock that increases by 20% or decreases by 10%.



Figure 3: Two-state model

## 1.3 Absence of arbitrage opportunities and valuation

Pure arbitrage theory is set in a perfectly competitive market for financial assets. At a given market price p, a stakeholder can buy or sell any quantity of assets. Moreover, there are no limits on allowable short positions. Thus,

the theory does not account for possible transaction costs, prohibitions on short sales, or limitations on buying or selling.<sup>2</sup>

An opportunity for arbitrage is a possibility to realize non-negative profits in all states of nature, today or in the future, with a strictly positive profit in at least one state.<sup>3</sup> In the absence of constraints on transactions, this leads to the following definition:

#### **Definition 1** Arbitrage opportunity

An opportunity for arbitrage in markets  $(\mathcal{E}, \tilde{a}, p)$  is a portfolio z such that:

$$\sum_{1}^{K} z_k a_k(e) \ge 0 \quad \forall \ e \ \text{ and } \ z'p \le 0,$$

(or, equivalently:  $z'\tilde{a} \ge 0$  and  $z'p \le 0$ ) with at least one strict inequality among these E + 1 inequalities.<sup>4</sup>

If there are no limits on the quantities exchanged, an opportunity for arbitrage cannot last, since operators will have an interest in exploiting it infinitely. This leads us to consider markets in which there are no opportunities for arbitrage.

A direct consequence of the *absence of opportunities for arbitrage* is that the value of a portfolio depends only on the payoff it generates. Indeed, assume that two portfolios,  $z^1$  and  $z^2$ , generate the same payoffs but do not have the same value. Then, for example, if:

$$p'z^1 < p'z^2$$

portfolio  $z^1 - z^2$  constitutes an opportunity for arbitrage.

 $<sup>^{2}</sup>$ In practice, purchase and sales prices differ to varying degrees, the difference being a *bid-ask spread*. Exercise 4 deals with a simple example of arbitrage relationships in this circumstance.

 $<sup>^{3}</sup>$ Two slightly different notions of arbitrage opportunity may be considered, depending on whether the profit is immediate or deferred. The relationships between these notions are examined in Exercise 6 in Chapter 3.

<sup>&</sup>lt;sup>4</sup>We adopt the following conventions for vector notation:  $z \ge 0$  means that each element of z is positive or nil, z > 0 is equivalent to  $z \ge 0$ , except that at least one component is strictly positive. Finally, z >> 0 indicates that all elements are strictly positive.

Valuation by arbitrage follows from this remark: A security j with revenue  $\tilde{a}_j$  is "replicated" by a portfolio z comprising other securities if the payoff yielded by z is identical to that from j in all states:

$$a_j(e) = \sum_{k \neq j} z_k a_k(e) \; \forall \, e.$$

In this case the security is said to be *redundant*. The price of security j must be equal to the value of the portfolio that replicates it, so as to eliminate any opportunity for arbitrage. We thus obtain a relationship between the price of the replicated security and those of the other securities in the replicating portfolio.

#### Example 2

#### Price of a call option in a two-state model

Let us return to the first example with two states of nature, one risk-free investment, one stock and a call option on the stock at time 1 with strike price K. The option can be replicated with the stock and the risk-free asset. Consider the more interesting case in which the option is only exercised if the stock price is high, so that

$$(1+h)S > K > (1+b)S.$$

The income yielded by the option is then:

$$\begin{cases} (1+h) S - K & \text{in state } h, \\ 0 & \text{in state } b. \end{cases}$$

For a portfolio z consisting of the risk-free asset and the stock to replicate the option, it must satisfy:

$$z_1(1+r) + z_2(1+h)S = (1+h)S - K,$$
  

$$z_1(1+r) + z_2(1+b)S = 0.$$

Since this system of equations has the solution:

$$z_1 = -\frac{(1+b)[(1+h)S - K]}{(1+r)(h-b)}, \ z_2 = \frac{(1+h)S - K}{(h-b)S},$$

the price of the option, C, must be equal to  $z_1 + z_2 S$ , so that:

$$C = \frac{r-b}{(1+r)(h-b)} \left[ S(1+h) - K \right].$$
(2.1)

Notice that this reasoning is valid on condition that the underlying security is available and can be traded without limitation, without transactions nor storage costs. If we are dealing with options on wheat prices, replication involves storage costs and is asymmetrical in terms of buying and selling, as we saw in the previous chapter.

#### Valuation using state prices

There exists a useful tool for pricing all portfolio payoffs without explicitly referring to the composition of the portfolio. These are state prices, which are frequently used—especially in the context of dynamic valuation. State prices play a role comparable to the discount factors used in intertemporal analysis without uncertainty. Just as discount factors allow comparison of revenues at different times, state prices allow the comparison of revenues between different states of nature. The state price associated with a state e has a direct interpretation when a specific security, known as an Arrow-Debreu security or a contingent security, is traded. An Arrow-Debreu security associated with a state of nature e yields one dollar if e occurs, and nothing otherwise. If such a security exists, its price is the state price associated with e: It is the price today of one dollar tomorrow in state e. Thus, it is a contingent discount factor. Even if there are no Arrow-Debreu securities, state prices can be defined whenever there are no arbitrage opportunities.

#### Theorem 2.1 State Prices

There are no arbitrage opportunities on markets  $(\mathcal{E}, \tilde{a}, p)$  if, and only if, there

exists a vector  $q = [q(e)]_{e \in \mathcal{E}}$  of strictly positive elements, such that

$$p_k = \sum_e q(e)a_k(e) \quad \forall k.$$
(2.2)

Vector q is called a vector of state prices.

Corollary 1 Discounting with state prices

Let q be a vector of state prices and z a portfolio. Then:

$$p'z = \sum_{e} q(e)c_z(e).$$
(2.3)

1		

According to this corollary,

The price today of a portfolio is equal to the sum of the portfolio incomes discounted by the state prices.

State prices allow the comparison of revenues across various states of nature. They make it possible to find the value today of any *replicable* income. A contingent income vector (c(e)) is said to be<sup>5</sup> replicable or spanned if there is a portfolio z that generates exactly the same payoff in each state  $e: c(e) = c_z(e)$ . Equivalently the vector is a linear combination of revenues generated by existing assets. Expression (2.3) can be restated as saying : The price payable today to obtain a replicable income vector tomorrow is equal to the sum of these incomes discounted by the state prices.

The qualifier *replicable* is very important here. The terminology "state prices" may be confusing, and sometimes leads to a misguided application of (2.3) for computing the "value" of an income vector (c(e)) that is not replicable: This point will be clarified when we distinguish between complete and incomplete markets.

Corollary 2 State prices and Arrow-Debreu prices

<sup>&</sup>lt;sup>5</sup>Note the terminology is identical to that used for a redundant asset: the payoff of a redundant asset is spanned by a portfolio composed of other assets.

Let q be a vector of state prices. If there exists a portfolio yielding 1 dollar in state e and 0 dollars otherwise, its price is equal to q(e).

Corollary 2 assumes that the Arrow-Debreu security contingent on e can be replicated by a portfolio. In this case, q(e) is unique, of course, and can be interpreted as the price to be paid today to obtain 1 dollar in state e. Otherwise, this interpretation is false and, furthermore q(e) is not uniquely determined. As an example of such a situation, let there be three states of nature and two securities, with:

$$\tilde{a} = \left[ \begin{array}{rrr} 1 & 1 & 0 \\ 1 & 1 & -1 \end{array} \right].$$

The price of the Arrow-Debreu security associated with state 1 cannot be determined by arbitrage. There are an infinity of state prices, q(1), satisfying the property in Theorem 1.

#### Proof of Theorem 1

The demonstration relies on a strong version of the Farkas lemma—see Gale, (1960, p.49) for example:

**Lemma**: Let A be a real  $K \times L$  matrix. Then only one of the following properties obtains:

- 1. There exists one solution x >> 0 to the equation Ax = 0.
- 2. There is one solution, z, in  $\mathbb{R}^{K}$  to the inequality z'A > 0.

Let  $A = (\tilde{a}, -p)$  be a  $K \times (E+1)$  matrix created by horizontally stacking  $\tilde{a}$  and -p. If (2) from the lemma obtains, there exists a portfolio z, such that  $z'\tilde{a} \ge 0$  and  $p'z \le 0$  with some strict inequality. This contradicts the absence of opportunities for arbitrage. Thus, (1) obtains, and the last element of x can be set equal to 1 without loss of generality. Denoting by q the vector of the E first elements, this gives:

$$\tilde{a}q - p = 0,$$

which proves (2.2). The converse is obvious.

#### 1.4 Complete markets

The notion of *complete markets* is useful for a thorough understanding of the fundamental limitations of valuation by arbitrage. Markets are complete when all revenue configurations are replicable through some portfolio.

#### **Definition 2** Complete markets

Given  $(\mathcal{E}, \tilde{a})$ , markets are complete if, for all  $\tilde{c} = [c(e)]_{e \in \mathcal{E}}$ , there exists a portfolio z such that  $\tilde{c} = \tilde{c}_z$  i.e.:

$$c(e) = \sum_{1}^{K} z_k \tilde{a}_k(e) \quad \forall e.$$

For markets to be complete, there must be at least as many securities as there are states of nature. It there is one security contingent on each state of nature, markets are clearly complete: The income configuration  $\tilde{c}$ is obtained by buying c(e) units of each security contingent on state e. The cost of  $\tilde{c}$  is thus simply equal to  $\sum_{e} q(e)c(e)$ . Indeed, if markets are complete, everything transpires as if such a full system of contingent securities existed, since there exists a portfolio yielding 1 dollar in state e and 0 otherwise. If we denote the value of this portfolio by q(e), we are back to the previous case. Rather than working with the original securities we can, by a linear transformation, revert to a complete system of contingent securities. The following property, the demonstration of which is left to the reader, characterizes complete markets:

**Theorem 2.2** Given  $(\mathcal{E}, \tilde{a}, p)$ :

(1) Markets are complete if and only if the rank of  $\tilde{a}$  is E.

(2) Complete markets without opportunities for arbitrage are associated with a unique vector of state prices q, and all future income configurations  $\tilde{c}$  have a present value given by the discounting formula:

$$\sum_{e} q(e)c(e).$$

In practice, eliminating possibly redundant securities, markets are complete if there are exactly as many linearly independent securities as there are states of nature. In that case, the square matrix  $\tilde{a}$  has an inverse, the vector of state prices, if any, is given by  $q = \tilde{a}^{-1}p$ . Thus, the condition of absence of opportunities for arbitrage is simply written:

$$q = \tilde{a}^{-1}p >> 0.$$

All financial instruments whose payoffs can be written as a function defined on that state space can thus be valued with state prices.

#### Example 3

Let us return to Example 2, in which there are two states of nature h or b, that determine the stock price increase (b < h). Markets are complete and the state prices q(h) and q(b), if any, satisfy:

$$1 = q(h)(1+r) + q(b)(1+r)$$
  

$$S = q(h)S(1+h) + q(b)S(1+b),$$

whence:

$$q(h) = \frac{r-b}{(1+r)(h-b)}, \quad q(b) = \frac{h-r}{(1+r)(h-b)}$$

The condition that state prices be positive is thus b < r < h (which we could easily have established by reasoning directly).

The price C of a call option of strike price K is simply:

$$C = q(h) \left[ S(1+h) - K \right] = \frac{r-b}{(1+r)(h-b)} \left[ S(1+h) - K \right].$$

Here, again, we find Equation (2.1), as in Example 2.

#### 1.5 Risk-adjusted probability

When there exists a risk-free security, the pricing formula in terms of state prices is interpreted as a mathematical expectation with respect to a probability measure calculated by normalizing the state prices. Such a distribution is called *risk-adjusted*. This formulation is very popular in finance, in particular in dynamic settings (see the end of this chapter).

By definition, a risk-free security yields the same payoff in all possible states at time 1: 1 + r dollars for 1 dollar invested at time 0, where r is the risk-free interest rate. When q is a vector of state prices, Equation (2.2) applied to the risk-free security yields:

$$1 = \sum_{e} q(e)(1+r).$$

Since state prices are non-negative, we can define a probability distribution  $\bar{\pi}$  on  $\mathcal{E}$  by:

$$\bar{\pi}(e) = q(e)(1+r).$$

Equation 2.2 then becomes:

$$p_k = \frac{1}{1+r} \sum_e \bar{\pi}(e) a_k(e).$$

Anticipating on further chapters, the revenue per dollar invested on security k,  $\tilde{a}_k/p_k$ , is called the (gross) *return* to security k. Note that the return to the risk free security is 1 + r. This immediately yields:

#### Corollary 3 Risk-adjusted probability

If there exists a risk-free security, there are no arbitrage opportunities if and only if there exists a probability distribution  $\bar{\pi}$  on  $\mathcal{E}$ , with positive probability on each state, such that the price of any security is equal to its discounted expected payoff:

$$p_k = \frac{1}{1+r} \sum_e \bar{\pi}(e) a_k(e) \quad \forall \, k,$$

where r is the risk-free rate.

Under this probability, the expected returns to securities are all equal:

$$\sum_{e} \bar{\pi}(e) \frac{a_k(e)}{p_k} = 1 + r \quad \forall \, k$$

By construction, the risk-adjusted probability is simply a system of normalized prices. The terminology may be confusing. There is no immediate relationship with the probabilities of occurrence of the various states of nature—these have not yet been defined!

# 2 Intertemporal arbitrage

The preceding results extend directly to an intertemporal framework with several periods. First, we specify the time structure.

## 2.1 Time structure

There are several periods with a final date T: t = 0, ..., T. At time T, all securities are sold at given exogenous prices. Transactions may occur, dividends be distributed, and information become available at discrete times intervening between 0 and T. To formalize this situation, we let there be a set of states of nature on each date from 1 to T.

As previously, a state of nature at a time t provides a full description of the economic environment at time t, i.e. all known factors that may impact on securities prices and their future dividends at time t. Past events are assumed not to be forgotten. In other words, knowledge of the state at time t includes knowledge of the states through which the economy passed since time 0. In practice,  $e_t$  is written as  $e_t = (e_{t-1}, \varepsilon_t)$ , where  $\varepsilon_t$  stands for the possible shocks occuring at date t. This structure is represented by a *tree* (using graph terminology). Figure 4 represents a model, called binomial, in which a state is followed by two possible states over four periods. We write  $e_t < e_{t+1}$  if state  $e_t$  precedes  $e_{t+1}$ , and we say that  $e_{t+1}$  is a successor to



Figure 4: A binomial tree structure

state  $e_t$ . A state can only have one predecessor, but typically has several successors. Seen from date 0, the set of all possible states at time T can be very large. For example, if each state has two successors, there will be  $2^T$  final states—as time elapses information becomes more focused, and the number of possible future final states is divided by two at each date. More generally, the set of all paths emanating from a state  $e_t$  until T provides the full range of possible developments. We denote by  $\mathcal{E}_t$  the set of all possible states at time t and  $\mathcal{E} = \bigcup \mathcal{E}_t$  the set of all states. This latter set plays an analogous role to the set of states  $\mathcal{E}$  (same notation) in the two-period model from the previous section.

#### 2.2 Instantaneous arbitrage

First, consider only portfolios constituted at time 0 and held until time T with no modification during the intervening dates.

The payoffs procured from the securities are described by extending the conventions of the two-period model. At all dates t, before the end of times, t < T, the income, in terms of the numeraire, paid to the owner of one unit of security k is  $d_k(e_t)$  in state  $e_t$ . For example, for a bond,  $d_k(e_t)$  is equal to the coupon paid in state  $e_t$ , possibly increased by reimbursement of the capital, if it matures before T. To account for assets representing claims that extend beyond time T, the income served at the last date is denoted  $a_k(e_T)$ : it includes the (exogenously given) resale price at the last date. Thus, in the case of a stock,  $d_k(e_t)$  represents the dividend paid out to stockholders in any state  $e_t$  prior to liquidation at time T, and  $a_k(e_T)$  is the sum of the dividend and the resale price in any state at time T. We sometimes designate the payoffs a and d with the generic term 'dividend', even though the terminology is not appropriate for a, and only valid for d in the case of stocks (for bonds, coupons would be the relevant terminology, at least before maturity).

By an immediate application of the absence of arbitrage opportunities, and in particular of (2.2) from Theorem 1, to all possible states at all times  $\mathcal{E}$ , there exist strictly positive state prices,  $q = [q(e_t)]$ , such that

$$p_k(e_0) = \sum_{e_t, t=1, \dots, T-1} q(e_t) d_k(e_t) + \sum_{e_T} q(e_T) a_k(e_T).$$
(2.4)

The intertemporal structure allows futures markets open at time 0 to be accommodated. For example, consider a futures market for time t for a fixedincome bond that yields a sure income stream, i.e. that is only a function of time, denoted d(t). This bond can be bought or sold for delivery at time t, at a price f agreed upon today and payable at t. Purchase of a bond on the futures market for t does not cost anything before that date and yields a revenue of -f, regardless of the states at time t and  $d(\tau)$  for all states at the times  $\tau$  subsequent to t. Thus, we have:

$$f\sum_{e_t\in\mathcal{E}_t}q(e_t)=\sum_{\tau>t}\left[\sum_{e_\tau\in\mathcal{E}_\tau}q(e_\tau)\right]d(\tau).$$

The absence of opportunities for arbitrage provides a condition for consistency of futures prices with spot prices. Moreover, introducing futures markets may contribute to making markets complete. However, the potential is limited since, as in the example above, futures markets are generally unconditional. Transactions are assumed to occur at time t, whatever state of nature  $e_t$  prevails at that time. Nonetheless, we have seen in Chapter 1 how, in the case of certainty, futures transactions are *replicable* by cash-andcarry type strategies, i.e. by interventions on spot markets carried out today and at the term. This type of argument can be generalized to the case of uncertainty: the existence of spot markets in the future, *provided that the prices to be established on them are fully anticipated*, allow markets to be *dynamically* completed.

#### 2.3 Dynamic arbitrage

Assume that the *only* markets in place are spot markets for securities, which are open at all times.<sup>6</sup>

Thus, the *data* are:

- a succession of states of nature described by a tree,
- securities  $k = 1, \ldots, K$ ,

- the sequence of dividends per unit of security k,  $d_k(e_t)$  in state  $e_t$ , at all intermediate dates, and a final payoff  $a_k(e_T)$  for the last date T,

- the sequence of security k spot prices  $p_k(e_t)$ . By convention,  $p_k(e_t)$  is the spot price of the security purchased in the state  $e_t$  after distribution of the dividend  $d_k(e_t)$ , for all intermediate dates. By convention, as in the static model, the payoff  $a_k(e_T)$ at the final date includes the resale price of the security, so that  $p_k(e_T)$  can be set to zero.

This information is assumed known by the participants. This means in particular that the spot prices of the securities,  $p_k(e_t)$ , are perfectly foreseen, or, equivalently, that the prices of the securities are part of the definition of  $e_t$ .

The existence of spot markets for securities that are open in all states of nature, and the assumption of *perfect foresight on prices* on these markets provide the framework for valuation by dynamic arbitrage. It is the most commonly used, both for discrete time (as here) and for continuous time. A portfolio constituted today is not necessarily maintained unaltered until time T. In particular, a stochastic revenue stream may be replicated with a program of acquisitions and sales during the intervening periods on the markets that will open, called a *dynamic portfolio strategy*.

<sup>&</sup>lt;sup>6</sup>The introduction of futures markets in parallel to these spot markets would seriously complicate the notation and, to a lesser extent, the analysis. In the particular case of dynamically complete markets, introduced further on, all futures transactions on an existing underlying security can be replicated by a sequence of cash-and-carry type operations on the spot market. Futures markets are thus redundant.

**Definition 3** A portfolio strategy z defines the portfolio,  $z(e_t) = [z_k(e_t)]$ , held in each state  $e_t$  after the transactions. Everything is liquidated at T,  $z(e_T) = 0$  for all final states  $e_T$ .

The value of the strategy z at time 0 is that of the initial portfolio:

$$p(e_0)'z(e_0) = \sum_{k=1}^{K} z_k(e_0)p_k(e_0),$$

and the income generated in a state  $e_t$  at date t, t = 1, ..., T - 1, is given by:

$$c_z(e_t) = \sum_{k=1}^{K} \left[ z_k(e_{t-1}) - z_k(e_t) \right] p_k(e_t) + \sum_{k=1}^{K} z_k(e_{t-1}) d_k(e_t),$$

and at date T:

$$c_z(e_T) = \sum_{k=1}^{K} [z_k(e_{T-1})] a_k(e_T)$$

where  $e_{t-1}$  is the unique predecessor of  $e_t$ ,

#### Example 4

Some strategies, called *elementary*, are very simple. They involve only intervening if a given state  $e_t$  occurs, and liquidating the purchased portfolio on the following date. If the portfolio consists of one unit of security k, the strategy viewed from time 0 consists of two opposing operations on the spot market for security k: At time t, if the state is  $e_t$ , one unit of the security is bought, and it is then resold on the following date whatever happens.

More generally, an elementary strategy is characterized by a vector  $\theta(e_t)$ in  $\mathbb{R}^K$  which describes the composition of the portfolio, and is written:

$$z(e_t) = \theta(e_t), \quad z(e_s) = 0 \text{ for } e_s \neq e_t.$$

The income generated by this strategy is:<sup>7</sup>

<sup>&</sup>lt;sup>7</sup>The formula is valid for t < T - 1. When t = T - 1, the income generated by the

$$c(e_t) = -\sum_{k=1}^{K} p_k(e_t) \theta_k(e_t), \qquad (2.5)$$

$$c(e_{t+1}) = \sum_{k=1}^{K} [p_k(e_{t+1}) + d_k(e_{t+1})] \theta_k(e_t) \text{ if } e_{t+1} > e_t \qquad (2.6)$$

$$c(e_s) = 0$$
 for any other state. (2.7)

An arbitrage opportunity is a strategy with an initial value that is negative or nil and that generates non-negative revenues at all times and in all future states, with at least one strict inequality.

**Definition 4** An opportunity for arbitrage is a strategy z such that  $c_z(e_t) \ge 0$  for all  $e_t$ , all t > 0, and  $p(e_0)'z(e_0) \le 0$ , with at least one strict inequality.

Intuitively, a short-term arbitrage opportunity in state  $e_t$ , between t and t + 1, should translate into an intertemporal opportunity for arbitrage. In addition, according to the principles of static arbitrage, the absence of opportunities for arbitrage in the short term implies the existence of state prices for the direct successors of  $e_t$ ,  $e_{t+1}$ . We shall demonstrate that this is sufficient for constructing state prices as of time 0. A necessary and sufficient condition for the absence of opportunities for intertemporal arbitrage is the absence of short term arbitrage opportunities in all states of nature.

#### Theorem 2.3

Assume an economy in which the only markets are spot markets and in which prices are perfectly anticipated conditionally on the states of nature.

1. The three following properties are equivalent:

strategy at date T is

$$c(e_T) = \sum_{k=1}^{K} a_k(e_T) \theta_k(e_{T-1}).$$

- (a) There are no opportunities for arbitrage.
- (b) There exists a vector  $q = [q(e_t), e_t \in \mathcal{E}]$  of strictly positive elements such that, for any strategy z:

$$p(e_0)'z(e_0) = \sum_{e_t, t=1,\dots,T} q(e_t)c_z(e_t).$$
(2.8)

- q is called the vector of state prices discounted at time 0.
- (c) For any state  $e_t$ ,  $t \ge 0$ , there exist a vector  $[q(e_{t+1}|e_t), e_{t+1} \in \mathcal{E}]$ , of strictly positive prices for all direct successors of  $e_t$ , such that, for all k:

$$p_k(e_t) = \sum_{e_{t+1}|e_{t+1}>e_t} q(e_{t+1}|e_t)[p_k(e_{t+1}) + d_k(e_{t+1})],$$
  
$$t = 0, \dots, T - 2 \qquad (2.9)$$
  
$$p_k(e_{T-1}) = \sum_{e_T|e_T>e_{T-1}} q(e_T|e_{T-1})a_k(e_T).$$

The vector  $[q(e_{t+1}|e_t)]$  gives prices for the successor states of  $e_t$  discounted in  $e_t$ .

2. Given discounted state prices  $[q(e_{t+1}|e_t)]$  for all  $e_t$ , the prices defined by:

$$q(e_{t+1}) = q(e_1|e_0) \dots q(e_{t+1}|e_t), \qquad (2.10)$$

where  $(e_0, e_1, \ldots, e_{t+1})$  is the unique path from  $e_0$  to  $e_{t+1}$ , are state prices.

Conversely, given state prices  $[q(e_t)]$ , the formula

$$q(e_{t+1}|e_t) = \frac{q(e_{t+1})}{q(e_t)}$$

defines prices for the successor states of  $e_t$  for all  $e_t$ .

The interpretation of (2.8) is the same as in a two-period model :

the value of a portfolio strategy in  $e_0$  is equal to the discounted value of the revenues it generates.

Similarly, the discounted state prices at some date knowing the state  $e_t$  link the price of a security with the income it generates at time t + 1. According to (2.9),

the price of security k in state  $e_t$  is equal to the discounted value, measured with the discounted state prices in  $e_t$ , of the income (dividends + resale price) it generates at time t + 1.

#### Proof of Theorem 3

1) (a)  $\Rightarrow$  (c) This results from applying Theorem 1 to the elementary strategies of state  $e_t$ .

(c)  $\Rightarrow$  (b) Let us define the prices  $[q(e_t)]$  by (2.10). We multiply (2.9) by  $q(e_t)\theta_k(e_t)$  and sum over k. This yields:

$$q(e_t) \left[ p(e_t)'\theta(e_t) \right] = \sum_{e_{t+1}|e_{t+1}>e_t} q(e_{t+1}) \left[ p(e_{t+1}) + d(e_{t+1}) \right]' \theta(e_t),$$
  
$$t = 0, \dots, T-2$$
  
$$q(e_{T-1}) \left[ p(e_{T-1})'\theta(e_{T-1}) \right] = \sum_{e_T|e_T>e_{T-1}} q(e_T)a(e_T)'\theta(e_{T-1}).$$

Using (2.5), we see that (2.8) obtains for all elementary strategies. Observe that any portfolio strategy z is a sum of elementary strategies  $\theta(e_{\tau})$ , setting

$$\theta(e_0) = z(e_0) \quad \theta(e_t) = z(e_t) - z(e_{t-1}) \text{ for } t > 0.$$

By linear combination, (2.8) is true of all portfolio strategies. This also demonstrates that q is a system of state prices.

(b)  $\Rightarrow$  (a) directly.

The first part of 2 has already been demonstrated. Conversely, applying (2.8) to elementary strategies associated with state  $e_t$ , dividing by  $q(e_t)$ , yields (2.9).

Equation (2.8) allows any revenue stream that can be replicated by a portfolio strategy to be valued. Drawing on the static model, this leads to the introduction of the notion of *dynamically complete markets*. In such markets,

any income stream can be replicated by a portfolio strategy. Consequently, (2.8) permits a rigorous interpretation of state prices, on one hand, and of the valuation of any financial instrument, on the other.

**Definition 5** Markets are dynamically complete if any income stream  $(c(e_t), t \ge 1)$ , can be replicated by a portfolio strategy: for any stream c, there exists z such that:

$$c(e_t) = c_z(e_t) \quad \forall e_t, t = 1, \dots, T.$$

Let us interpret state prices in the case of complete markets. According to (2.8), the cost of a strategy that yields one dollar in state  $e_t$  and nil in any other state is necessarily equal to  $q(e_t)$ . Consequently,  $q(e_t)$  is the price to be paid at t = 0 to obtain one dollar in state  $e_t$ . It is strictly positive and uniquely defined. When markets are incomplete, this interpretation is only valid if such a strategy exists. The most important practical point is that markets are dynamically complete whenever there are enough securities in each state to generate any vector of state-contingent revenues for the *immediately* succeeding states. Compared with the static perspective at the initial date, which requires considering all paths that the economy might follow, the use of dynamic strategies allows a considerable reduction in the number of securities required for complete markets. When it is a matter of a derivative written on an underlying security, the relevant states correspond to the various prices of the security, which are naturally organized into a tree describing the possible evolution of this price. If markets are complete in each period—and this condition imposes a strong constraint on the choice of the tree—it is technically simple to assign a price to each new derivative for each price stream. We shall illustrate these techniques with the binomial model of Cox, Ross and Rubinstein (1979).

#### Valuing an option in the binomial model

We take up Example 2 extended to several periods. In each period, a risk-free security with a constant return of r between two successive dates

and the risky stock can be traded. A state  $e_t$  is followed by two successors,  $(e_t, h)$  and  $(e_t, b)$ , corresponding to stock price growth rates of h and b, respectively. In other words, if  $S(e_t)$  is the stock price in state  $e_t$ , its price will be  $S(e_t)(1+h)$  in state  $(e_t, h)$  or  $S(e_t)(1+b)$  in state  $(e_t, b)$ .

Consider a derivative on this stock, say a European option maturing at T with a strike price of K. From the perspective of time 0 there are a large number of final states. However, markets are dynamically complete: Since each state is followed by two successors, the risky stock and the risk-free security suffice to complete the markets (their payoffs are never proportional since b and h differ). To value the option using (2.8), we need to evaluate the state prices. For this, it is convenient to first compute the state prices between two successive dates, and then apply (2.10).

• Calculation of the state prices  $q(e_{t+1}|e_t)$ 

We apply (2.9) to the two assets in  $e_t$ , knowing that the two following states are characterized by h and b:

$$1 = q(h|e_t)(1+r) + q(b|e_t)(1+r),$$
  

$$S(e_t) = q(h|e_t)S(e_t)(1+h) + q(b|e_t)S(e_t)(1+b).$$

Thus, it follows that state prices are independent of  $e_t$ , and are given by the same expression as in the two periods Example 2:

$$q(h) = \frac{r-b}{(1+r)(h-b)}, \quad q(b) = \frac{h-r}{(1+r)(h-b)}.$$

#### • Calculation of the state prices q discounted in 0.

A state  $e_t$  is characterized by the succession of growth rates, high or low, realized from date 0 up to t. The state price, which is the product of the intermediate prices, is equal to  $q(h)^i q(b)^{t-i}$  if there were exactly *i* times *h* and t - i times *b* between 0 and *t*. Consequently, it is independent of the order in which the jumps occurred.

• The price at t = 0 of an option maturing at T and with a strike price of K.

The option does not distribute intermediate dividends, and at time T it

pays out  $[S(1+h)^i(1+b)^{T-i}-K]^+$  (we denote  $\alpha^+ = \max(\alpha, 0)$ ) if there were *i* high yields between 0 and *T*. Setting  $\bar{\pi} = (r-b)/(h-b)$  and grouping all the states in which the yield was high *i* times, we obtain:

$$C = \frac{1}{(1+r)^T} \sum_{i=0,\dots,T} \frac{T!}{(T-i)!i!} \bar{\pi}^i (1-\bar{\pi})^{T-i} \left[ S(1+h)^i (1+b)^{T-i} - K \right]^+$$

 $\bar{\pi}$  can be interpreted as a probability of occurrence of h, since it is between zero and one.

In the Cox-Ross-Rubinstein model the option price is equal to the expectation, given the (risk-adjusted) probability  $\bar{\pi}$ , of its final value discounted at the risk-free rate.

This result can be generalized when the risk-free rate varies with the state of nature (see the following section).

• Hedging strategies

The use of state prices allows the option price to be computed very easily. Often a financial institution also wishes to calculate the portfolio strategy that replicates the option. Indeed, having sold the option and unwilling to assume the associated risk, a replicating portfolio strategy serves as a hedge: it exactly gives the payoffs that the institution is required to pay to the option holders. The following algorithm accomplishes this:

- 1. Starting from the "end" of the tree, in any state  $e_{T-1}$  preceding maturity, compute the value of the option  $C(e_{T-1})$  in these states and the portfolio  $z(e_{T-1})$  that replicates it.
- 2. For any state  $e_{T-2}$  at T-2, we know from step 1 the option price in the two successor states:  $C(e_{T-2}, h)$  and  $C(e_{T-2}, b)$ . The value of the option in  $e_{T-2}$  follows:

$$C(e_{T-2}) = q(h)C(e_{T-2}, h) + q(b)C(e_{T-2}, b),$$

This is also the value of the portfolio  $z(e_{T-2})$  that will be worth  $C(e_{T-2}, h)$  and  $C(e_{T-2}, b)$  in the two following states. This portfolio permits the purchase of the replicating portfolio in the two succeeding

states.

### 3. Continue "up" the tree in the same fashion until time 0.

This example illustrates two points:

- The usefulness of the valuation formula (2.8): if we only seek to value the option it is much simpler to compute the state prices than to replicate the option. The valuation however is only valid when such a replicate exists, which follows automatically if markets are dynamically complete;

- The flexibility provided by the algorithm : it allows to price and replicate all derivatives written on the stock. For example, consider an American option (recall that such an option can be exercised at various dates before maturity). In each state, compare the value of the option if it is not exercised with the profit generated by exercising it immediately: this simultaneously determines the value of the option as the greater of these two terms, and the optimal exercise strategy.

#### Example 5

#### Valuation of an option with a sliding strike price

The binomial tree allows the valuation of options with complex characteristics, known as "exotics." Here we propose the valuation of an option on stocks with the following features:

• The strike price is periodically redefined as a percentage of the stock price, unless it reaches a floor fixed at the time of issue.

• The option can be exercised at certain predetermined periods.

Figure 5 represents the possible evolution of the stock price over four periods, in which we have set h = 0.20 and 1 + b = 1/(1 + h). If the risk-free rate r is equal to 0.05, the risk-adjusted probability of the high state  $\bar{\pi}$  is equal to 0.591.

Consider an option with a strike price of 95 at time 0, revisable at t = 2, and whose new value will be the greater of 90 (the floor) or 95 per cent of the stock price. The option can only be exercised on even dates, t = 2, 4 and at the strike price determined two periods previously. Thus, 95 is the strike price that will prevail if it is exercised at time 2. At this time, the strike price (which is to be exercised at t = 4) is adjusted to 136.8 if S = 144, maintained at 95 if S = 100, and lowered to 90 if S = 69.4.



Figure 5: Binomial model

The calculation of the option value proceeds in several steps, starting from the end of the tree. The retention value of the option at a given date is defined as the value yielded by the option if it is not exercised immediately.

**Step 1**. We start at t = 2 in one of the three possible states.

Assuming that we keep the option, it can only be exercised at t = 4 and at a known strike price. Thus, computing the retention value is a simple matter.

This allows the exercise strategy to be determined at t = 2. To know whether it is preferable to exercise the option immediately or hold on to it, all that is needed is to compare the retention value with the profit yielded by exercising it now.

For example, assume that S = 144. The retention value is equal to

$$\bar{\pi}^2(207.4 - 136.8) + 2\bar{\pi}(1 - \bar{\pi})(144 - 136.8) = 28.14,$$

while exercising it immediately yields 144-95 = 49: The option is exercised! Similarly, we find that

- if S = 100: the retention value,  $\bar{\pi}^2(144 95) + 2\bar{\pi}(1 \bar{\pi})(100 95) = 19.53$ , is larger than the exercise value, 5, (the option is not exercised), and
- if S = 69.4: the retention value,  $\bar{\pi}^2(100 95) = 1.75$ , is larger than the exercise value, 0, (the option is not exercised),

respectively.

The option value at t = 2, before possibly being exercised, is the maximum of the two quantities, or 49 if S = 144, 19.53 if S = 100, and 1.75 if S = 69.4.

**Step 2**. We can now easily compute the option price at t = 0. It is the discounted sum, computed with the state prices, of its value at time 2:

$$\bar{\pi}^2 49 + 2\bar{\pi}(1-\bar{\pi})19.53 + (1-\bar{\pi})(1-\bar{\pi})1.75 = 26.984$$

#### 2.4 Probabilistic formulation: risk-adjusted probability

The notion of risk-adjusted probability introduced in the two-period model (section 1.5) is particularly useful in a dynamic framework. Indeed, after normalization, the formula (2.10) for constructing the state prices is transformed into the Bayesian formula for conditional probabilities.

Consider a *T*-period model. Assume that there exists a short-term risk-free security at all dates. Its return between t and t + 1, knowing the state  $e_t$ , is independent of the successor  $e_{t+1}$  of  $e_t$  and denoted by  $r(e_t)$ . Note that  $r(e_t)$ , which is the short term *risk-free* rate, depends upon the state  $e_t$  and

may vary over time: The security is risk-less only between two successive periods, once the state is known.

**Theorem 2.4** Assume an economy with a risk-free short-term security in each period. The three following properties are equivalent:

- 1. There are no opportunities for arbitrage,
- 2. There exists a strictly positive probability distribution  $\bar{\pi}$  for the tree, such that the value of any strategy is equal to the expectation under  $\bar{\pi}$  of the discounted sum at the risk-free rate for the future incomes it generates:

$$p(e_0)'z(e_0) = \sum_{e_t, t=1, \dots, T} E_{\bar{\pi}} \prod_{\tau=1}^t \frac{1}{1 + r(e_{\tau-1})} c_z(e_t),$$

3. There exists a strictly positive probability distribution for the transition from any state  $e_t$  to its successors  $\bar{\pi}(.|e_t)$  such that, for any k and  $e_t$ :

$$p_k(e_t) = \frac{1}{1+r(e_t)} E_{\bar{\pi}} \left[ p_k(e_{t+1}) + a_k(e_{t+1}) \mid e_t \right],$$
  

$$t = 0, \dots, T-2 \qquad (2.11)$$
  

$$p_k(e_{T-1}) = \frac{1}{1+r(e_{T-1})} E_{\bar{\pi}} \left[ a_k(e_T) \mid e_{T-1} \right].$$

-		
I		
l		

**Comments** As in the two-period model,  $\bar{\pi}$  is called the risk-adjusted probability. This probability is particularly useful in the dynamic model and in continuous time, which is obtained in the limit when the interval separating successive transactions is allowed to tend towards zero. In this situation powerful tools developed for stochastic processes can be used.

For a security k that does not distribute any dividends before the final period, assuming the interest rate is constant and equal to r, Equation (2.11) is often written as

$$(1+r)^{-t}p_k(e_t) = E_{\bar{\pi}}\left[(1+r)^{-t-1}p_k(e_{t+1}) \mid e_t\right].$$

The term  $(1+r)^{-t}p_k(e_t)$  is the security price in state  $e_t$  counted in terms of dollars at time 0, for short the discounted price. Thus, the above expression says that the expectation at time t + 1 of the discounted security price, conditional on the state at date t, is equal to its value in state  $e_t$ . The risk of loss compensates for the chance of profit in mathematical expectation. In mathematical terms, the discounted security price is a martingale.<sup>8</sup>

#### Proof of Theorem 4

All that is required is Theorem 3 and to define  $\bar{\pi}$  from the state prices  $[q(e_t)]$  with

$$\bar{\pi}(e_t) = \frac{q(e_t)}{\sum_{\text{states at time } t} q(e)}$$

and the transition probabilities from the transition prices, with

$$\bar{\pi}(e_{t+1}|e_t) = q(e_{t+1}|e_t)[1+r(e_t)].$$

#### **Bibliographical Note**

Arbitrage theory was pioneered in the work of Ross (1976) and (1978), who understood the importance for valuing a security of being able to replicate it with existing assets. Its application to options can be found in Cox, Ross and Rubinstein (1979). Our presentation of the static version in this chapter resembles that in Varian (1987), though we also call on a strong version of the separation theorem developed in Gale (1960). Cox, Ingersoll and Ross (1985) developed the dynamic interpretation in discrete time. We deliberately ignored the technical difficulties associated with continuous time models in order to focus on the basic concepts and economic principles. The reader wishing to learn more on this subject is referred to Demange and

$$E_t x_{t+1} = x_t.$$

<sup>&</sup>lt;sup>8</sup>Mathematically, a stochastic process  $\tilde{x}_t$  is a martingale if the expectation of  $x_{t+1}$  conditional on the information available at time t, is equal to  $x_t$ :

Rochet (1992). Finally, Hull (1997) is a reference on valuing derivatives and the associated hedging issues.

Cox, J, S. Ross and M. Rubinstein (1979): "Option Pricing : A Simplified Approach", Journal of Financial Economics, 7, 229-263.

Cox, J., J. Ingersoll and S. Ross (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-408.

Demange, G., and J-C. Rochet (1992) : *Méthodes Mathématiques de la Finance*, Economica, Paris 2nd edition 1997.

Gale, D. (1960): *The Theory of Linear Economic Models*, McGraw-Hill, New-York.

Hull J.C. (1997): Options, futures, and other derivatives, Prentice Hall.

Ross, S. (1976): "Risk, Return, and Arbitrage," in I. Friend and J. Bicksler ed. *Risk and Return in Finance*, Ballinger, Cambridge.

Ross, S. (1976): "Options and Efficiency," *Quarterly Journal of Economics*, 90, 75-89.

Ross, S. (1978): "A Simple Approach to the Valuation of Risky Streams," *Journal of Business*, 51, 3, 453-475.

Varian, H.R. (1987): "The Arbitrage Principle in Financial Economics," The Journal of Economic Perspectives, vol. 1, 2, 55-72

#### Exercises

#### **Exercise 2**: Options and complete markets

The notion of complete markets is essential to valuation by arbitrage. The following examples are designed to illustrate the notion that it may be possible to complete markets by authorizing the negotiation of options on pre-existing assets.

1. In a two-period model with three states of nature  $e_1$ ,  $e_2$ , and  $e_3$ , consider a single asset with payoffs  $\tilde{a}_1$  given by (4, 3, 1).

Show that the introduction of two call options with different strike prices on the asset allows the market to be completed.

How would this result be changed if the revenue  $a_1$  in state  $e_3$  were no longer 1, but 3? Explain your results.

2. Consider a second scenario in which the market is comprised of two assets. The states of nature in the second period are  $e_1$ ,  $e_2$ ,  $e_3$ , and  $e_4$ , and the revenue vectors are:

	$e_1$	$e_2$	$e_3$	$e_4$
$a_1$	1	1	2	2
$a_2$	1	2	1	2

Is it possible to complete this market by introducing options on  $a_1$  and  $a_2$ ? Explain your results.

3. Prove that it is possible to constitute a portfolio (or fund) with the two assets so that call options on this fund allow to complete the market.

Note : Question 1 illustrates the following, more general, result: Options on the securities can complete market only if the securities allow the different states of nature to be distinguished. Question 2 reveals that the converse is not trivial. Indeed, the condition that the income from the available securities allows a distinction to be made between the various states is necessary and sufficient for the existence of a *portfolio* of initial securities for which call (or put) options allow the market to be completed (see Ross, 1976).

**Exercise 3**: Consider a firm whose underlying value per share increases either by h or b between t = 0 and t = 1 and between t = 1 and t = 2, where h > b. At date 1 it pays out a dividend of d, so that if the share price at t = 0 is 1, the share price after the dividend is paid evolves as follows:

$$p(e_{1}) = (1 + h - d)$$

$$p(e_{1}) = (1 + h - d)$$

$$p(e_{4}) = (1 + h - d)(1 + h)$$

$$p(e_{5}) = (1 + b - d)(1 + h)$$

 $p(e_0) = 1$ 

$$p(e_2) = (1 + b - d)$$

$$p(e_6) = (1 + b - d)(1 + b)$$

$$t = 0$$

$$t = 1$$

$$t = 2$$

The risk-free interest rate is r.

1. Compute the state prices associated with  $e_1$  and  $e_2$ , at time t = 0. Verify that they are independent of the dividend. Under what condition are there no opportunities for arbitrage? Similarly, at  $e_1$  and  $e_2$ , compute the state prices of the two possible successor states.

2. Consider a call option that can be exercised at t = 2, with a strike price of K, and with (1 + h - d)(1 + b) < K < (1 + h - d)(1 + h). Find an expression for its price at t = 0.

3. Consider a call option that can be exercised at t = 1 before distribution of the dividend, or at t = 2 at the strike price K. Under what condition on d is it advantageous to exercise this option at t = 1?

4. Compute both option prices at t = 0 for:

$$h = 0.05, b = 0.01, r = 0.02, d = 0.04, K = 1.04.$$

#### **Exercise 4**: Arbitrage and transaction costs

The organization of markets creates operating costs that are supported by the participants. The purpose of this exercise is to examine some of the interactions between these costs and arbitrage operations.

Consider a two period model in which there are two states of nature and two securities. The first security is risk-free indexed with 0. To simplify, the interest rate is nil, so that one dollar invested in it yields one dollar in each state of nature. The other security yields a(1) in state 1 and a(2) in state 2, with a(2) > a(1) > 0. The purchase price of this security is denoted by  $p_1^+$ , may be greater than its sale price,  $p_1^-$ .

We wish to value another security, defined by a contingent revenue stream [b(1), b(2)] in the two states of nature, where b(2) > b(1). This security is exchanged on a market with no transaction costs: there is a single buy and sell price,  $p_2$ .

1. In the absence of opportunities for arbitrage, calculate the state prices  $q(e_1)$  and  $q(e_2)$  when  $p_1^+ = p_1^-$ . Use this to derive the price of security 2. Now assume that  $p_1^+ > p_1^-$ .

a) Determine the portfolio  $(z_0, z_1)$  of assets 0 and 1 that replicates the asset b. Verify that  $z_1 > 0$ .

b) In the absence of opportunities for arbitrage, determine upper and lower bounds  $\overline{p}_2$  and  $\underline{p}_2$  for the price of asset 2.

2. Here we examine how the Arrow-Debreu formula for asset valuation (Theorem 2) can be extended.

a) Calculate the bounds for the prices of the Arrow-Debreu assets:  $(q^+(e_i), q^-(e_i))$  for the asset yielding 1 in state *i*.

b) Assume that b(1) > 0. Show that the pricing formula:

$$p_2^+ = q^+(e_1)b(1) + q^+(e_2)b(2), p_2^- = q^-(e_1)b(1) + q^-(e_2)b(2),$$

yields upper and lower bounds for the asset price, but that these bounds can be improved upon.

# Chapter 4

# Portfolio choice

This chapter examines the portfolio choices of investors who allot a given amount of wealth to various securities. Determining how much wealth to invest is studied in subsequent chapters. "Rational" investors select a portfolio that best suits their objectives and their needs. Their demands for financial securities are derived from preferences represented with a von Neumann Morgenstern utility criterion, as discussed in the previous chapter. We are especially interested in how the selected portfolio is related to attitudes towards risk, (possibly risky) future non financial incomes, and assessment on securities payoffs.

Section 1 examines a particular case, referred to as the *mean-variance* analysis, that merits a detailed examination. Under some specifications, an investor ranks portfolios solely on the basis of the expectation and variance of their returns. Thus, given his budget constraint, he selects a portfolio that is *mean-variance efficient*, meaning that the expected return cannot be increased without also increasing the variance. An examination of all mean-variance efficient portfolios provides a first approach to the notion of *risk diversification*, the basis of many widely-used models in finance. When a risk-free security is available, the two "funds" theorem is obtained: meanvariance efficient portfolios are combinations of the risk-free security and a single portfolio of risky securities, analogous to a mutual fund. The specific composition of the risky fund is independent of the investors attitudes towards risk, as long as these investors have the same beliefs on the expectation and the variance of the securities returns. Thus the optimal composition of the risky fund represents an ideal diversification of risky investments. Attitudes vis-à-vis risk only come into play in the determination of the respective amounts invested in the risk-free security and the fund.

No such clear cut results hold in general: Section 2 analyses the demand for financial securities with an unrestricted von Neumann Morgenstern criterion. Finally, two specific cases, that of a quadratic utility function or of constant absolute risk aversion and normal returns, are studied in more details. They illustrate one of the themes underlying portfolio choice, *speculative* and *hedging* demand, and provide a link to mean-variance analysis.

# **1** Mean-variance efficient portfolios

This section presents the mean-variance framework. Whereas it can be linked to a von Neumann Morgenstern model (see the final section), this framework is used in practice for convenience. Indeed, it is simpler to work exclusively on the expectation and variance of a random variable rather than on its entire distribution.

#### The mean-variance hypothesis:

An investor ranks portfolios on the basis of the expectations and variances of their payoffs. The ranking is increasing in expectation and decreasing in variance.

Note that we do not fully specify the investor's attitude towards risk: In particular, we do not address how he arbitrates between the expectation and the variance of the payoff.

Under the mean-variance hypothesis, the investor selects a portfolio that is *mean-variance efficient*, meaning that the expected payoff cannot be increased without also increasing the variance. Of course, the expectation and the variance are evaluated by the investor, using his own beliefs.

As defined in the previous chapter, recall that a portfolio  $(z_*, z)$  with an

initial value of  $\omega_0 = p_* z_* + p' z$  yields a stochastic payoff in period 1:

$$\tilde{c} = z_* + \sum_{k=1}^{K} z_k \tilde{a}_k = z_* + z' \tilde{a}.$$

The expectation and the variance of the portfolio payoff are thus respectively given by  $z_* + Ez'\tilde{a}$  and  $\operatorname{var}(z'\tilde{a})$ . The set of mean-variance efficient portfolios is found by solving the following family of programs parameterized with a scalar M: <sup>1</sup>

$$\begin{cases} \min \operatorname{var}(z'\tilde{a}) \\ z_* + Ez'\tilde{a} \ge M \\ p_*z_* + p'z = \omega_0. \end{cases}$$

The parameter M is interpreted as the smallest expected payoff that the investor has set as a target. The solution to this problem is the portfolio(s) with the smallest variance meeting this target.

Thus, the mean-variance efficiency criterion does not determine a unique portfolio. Indeed, to derive the demand for securities, one needs to specify the investor's attitude towards risk, which in turn dictates the trade-off between the expectation and variance of the payoff. However, any solution to the above program—any efficient portfolio—features the key property of risk *diversification*, regardless of the level of risk aversion.

We first demonstrate that the analysis can be conducted on the basis of the *return* on each dollar invested. Then we characterize efficient portfolios in two stages, starting with the situation in which no risk-free security is available.

# 1.1 Portfolio composition and returns

The set of all mean-variance efficient portfolios is homogeneous of degree 1 in wealth  $\omega_0$ . Indeed, starting with a solution  $(z_*, z)$  to the program for the parameters  $(\omega_0, M)$ ,  $(\lambda z_*, \lambda z)$  solves the program for  $(\lambda \omega_0, \lambda M)$  for any positive  $\lambda$ . This naturally leads us to work per dollar invested, as is done in

<sup>&</sup>lt;sup>1</sup>The analysis conducted in this chapter is valid when the number of states of nature is infinite, provided that the variance  $var(\tilde{a})$  is finite.

finance. To this purpose, the *rate of return* of a security describes the payoff obtained from investing one dollar in the security, and the *composition* of a portfolio describes how much of each dollar is invested in each security (rather than the number of shares of each security in a portfolio). Formally, this is equivalent to a change of variables:

**Definition 1** The gross rate of return of a security is the stochastic payoff that one dollar invested in this security today pays at time 1:

$$\tilde{R}_k = \frac{\tilde{a}_k}{p_k}, \quad k = *, 1, .., K.$$

A portfolio composition is a vector  $(x_*, x)$ , such that  $x_* + \sum_{k=1}^{K} x_k = 1$ , where  $x_k$  is the fraction of the portfolio value invested in k.

Notice that the rate of return of a risk-free security is constant, given by  $R_* = (1 + r)$ . Frequently, the *net returns* equal to the gross returns minus 1 are used:  $\tilde{r}_k = \tilde{R}_k - 1$ . We often abbreviate 'rate of return' into 'return'.

Let us rewrite the investor's problem with these variables. The value of the portfolio  $(z_*, z)$  is  $\omega_0 = p_* z_* + p' z$ . If  $\omega_0$  is not nil, the portfolio composition is given by  $(x_*, x)$ :

$$x_* = \frac{p_* z_*}{\omega_0}, \quad x_k = \frac{p_k z_k}{\omega_0}, \quad k = 1, \dots, K.$$

Thus, the portfolio is characterized by its value and its composition. Moreover, the associated stochastic payoff satisfies:

$$\tilde{c} = z_* + z' \ \tilde{a} = \omega_0 (x_* R_* + x' \tilde{R}).$$

As for a single security, the gross return of a portfolio is equal to its random payoff per dollar invested. According to the expression above, the gross return of a portfolio whose composition is  $(x_*, x)$  is given by

$$x_*R_* + x'\tilde{R} = x_*R_* + \sum_{k=1}^K x_k\tilde{R}_k,$$

or, in words :

The return of a portfolio is the linear combination of the returns of the component securities weighted by their respective shares in the portfolio composition.

Letting  $\mathbf{1}_{K}$  represent the K-dimensional column vector consisting entirely of ones, mean-variance efficiency can be written in terms of the new variables:

**Definition 2** The portfolio with composition  $(x_*, x)$  is mean-variance efficient if it solves:

$$\min \operatorname{var}(x'\tilde{R}) \ s.t. \ \left\{ \begin{array}{rrr} x_*R_* + Ex'\tilde{R} & \geq & m \\ x_* + \mathbf{l}'_K x & = & 1. \end{array} \right.$$

for some value of m.

The parameter m is interpreted as the smallest expected return per dollar invested that the investor has set as a target, and the solution to this problem yields the portfolio(s) with the smallest variance meeting this target.

#### 1.2 Diversification

Assume that no risk free security is available. What is the best way to combine investments so as to minimize risks? The variance of a portfolio returns is simply expressed as a function of the variance-covariance matrix  $\Gamma$  of the returns of risky securities. This matrix is given by  $\Gamma = (\gamma_{hk})$  where  $\gamma_{hk}$  is the covariance between the returns of securities h and k. The variance

of the return of a portfolio of composition x is<sup>2</sup>

$$\operatorname{var}(x'\tilde{R}) = x'\Gamma x$$

Consider first the "degenerate" case, in which all securities have the same expected return,  $\overline{R}$ . In this case, the expected returns of all portfolios are identical and equal to that value. A portfolio is thus mean-variance efficient if and only if its variance is lowest. We are seeking x, the solution to

$$\min x' \,\Gamma x \text{ s.t. } \mathbf{1}'_K x = 1,$$

If  $2\mu$  is the Lagrange multiplier associated with the constraint  $\mathbf{1}'_{K}x = 1$ ,

$$x = \mu \Gamma^{-1} \mathbf{1}_K$$

where  $\mu$  is computed so as to satisfy the constraint.

Thus, there exists a unique optimal composition. For example, assume that all returns are independent and have the same variance : The variance-covariance matrix is proportional to the identity matrix. The optimal composition is then given by  $\mathbf{1}_K/K$ : The same amount, 1/K, should be invested in each security. This is *diversification*. Otherwise, without independence or identical variance, the optimal composition reflects differences between the variances of the security returns and their correlations.

In the general case in which expected returns differ across securities, it

```
^{2}x is a K-dimensional column vector and \Gamma is a (K \times K) matrix. We have
```

$$\gamma_{hk} = \sum_{e=1}^{E} \pi(e) \left[ \tilde{R}_h(e) - E\tilde{R}_h \right] \left[ (\tilde{R}_k(e) - E\tilde{R}_k \right].$$

and can write

$$\Gamma = E[(\tilde{R} - E\tilde{R})(\tilde{R} - E\tilde{R})'],$$

where the ' denotes transposition. The variance of the return of a portfolio composed of  $(x_*, x)$  is

$$\operatorname{var}(x'\tilde{R}) = Ex'(\tilde{R} - E\tilde{R})(\tilde{R} - E\tilde{R})'x$$

or  $x'\Gamma x$ . If we assume that there is no redundancy, the matrix  $\Gamma$  has an inverse. Otherwise, there would exist a non-nil vector x such that  $x'(\tilde{R} - E\tilde{R}) = 0$ , so that a portfolio of risky securities would replicate the risk-free security.

is necessary to arbitrate between the expectation and the variance of the return. For a given target on expected return, there exists a portfolio composition that minimizes risk. A graphical representation is helpful. Figure 7 plots the set of couples (standard error, expectation) associated with each possible portfolio in a  $(\sigma, m)$ -space. The couples associated with efficient compositions are on the frontier of that set, and constitute what is called the *efficiency frontier*.

#### **1.3** The efficiency frontier in the absence of a riskless security

Assume that there are at least two securities with different expected returns. The problem to be solved is written:

$$\min x' \Gamma x \text{ s.t.} \begin{cases} Ex'\tilde{R} = m \\ \mathbb{1}'_{K}x = 1, \end{cases}$$

for some values of m in  $\mathbb{R}$ . In contrast to the preceding formulation, the inequality on the expected return is replaced by an equality. For each target m, a solution is a portfolio with the smallest variance whose expected return is *exactly* m. Since there are two securities with distinct expected returns, the domain defined by the two constraints is not empty, regardless of the value of m (we are using the assumption that there are no limits on short sales). Imposing an equality constraint on the expected returns simplifies the math. Extending this solution to the formulation with inequalities, which is associated with the economic problem, is trivial.

Since the function to be minimized is a convex quadratic form, bounded below by zero, and the constraints are linear, the solution exists and is characterized by the first-order necessary and sufficient conditions. Letting  $2\lambda$  and  $2\mu$  respectively represent the multipliers for the expected return and the budget constraints, these conditions are:

$$\Gamma x = \lambda E \tilde{R} + \mu \mathbf{1}_K. \tag{4.1}$$

They give together with the two constraints a linear system of K + 1 equations in the K + 1 unknowns  $(x, \lambda, \mu)$ . To solve this system, x can be

expressed as a function of the multipliers from the first-order condition (4.1) because under the assumption of no redundancy the matrix  $\Gamma$  has an inverse. Plugging this expression into the constraints gives.

$$E\tilde{R}'\Gamma^{-1}E\tilde{R}\lambda + E\tilde{R}'\Gamma^{-1}\mathbf{1}_{K}\mu = m,$$
  
$$\mathbf{1}'_{K}\Gamma^{-1}E\tilde{R}\lambda + \mathbf{1}'_{K}\Gamma^{-1}\mathbf{1}_{K}\mu = 1.$$

This is a symmetric linear system with a strictly positive determinant. Inverting  $it^3$  yields

$$\lambda = \frac{dm-b}{\Delta}$$
 and  $\mu = \frac{-bm+c}{\Delta}$ .

The optimal portfolio is simply obtained by using again (4.1). We keep this calculation for later and first focus on finding the expression for the least variance as a function of m. For this, premultiply the first-order condition by x', yielding:

$$\sigma^2 = \lambda m + \mu,$$

where  $\sigma^2 = x' \Gamma x$  is the value of the variance of the efficient portfolio. Inserting the expressions obtained for  $\lambda$  and  $\mu$  yields

$$\sigma^2 = \frac{dm^2 - 2bm + c}{\Delta},$$

or

$$\left(\frac{\Delta}{d}\right)\sigma^2 - \left(m - \frac{b}{d}\right)^2 = \frac{\Delta}{d^2}.$$

This is the equation for a hyperbola in the (standard error, mean),  $(\sigma, m)$  plane. The standard error is lowest at m = b/d, where its value is  $1/\sqrt{d}$ . In terms of the initial problem, where we sought the portfolio with the smallest variance generating an expected return of at least m (we are reverting to the formulation with inequality rather than equality), only the section of the hyperbola in which the expected return exceeds b/d is of interest. For

<sup>&</sup>lt;sup>3</sup>We set  $d = \mathbf{1}'_K \Gamma^{-1} \mathbf{1}_K$ ,  $b = \mathbf{1}'_K \Gamma^{-1} E \tilde{R}$ ,  $c = E \tilde{R}' \Gamma^{-1} E \tilde{R}$ ,  $\Delta = dc - b^2$ . It is trivial to verify that  $\Delta$  is strictly positive provided there are two securities with different expected returns.



Figure 7: The two-fund theorem

all values of m below b/d, we wish to retain the least-variance portfolio. Any risky portfolio entails a level of risk equal to at least the minimum standard error  $1/\sqrt{d}$ : the non-diversifiable minimum risk. An investor who is prepared to accept a risk exceeding that minimal level can obtain a higher expected return.

# 1.4 Efficient portfolios: the case with a risk-free security

Now assume a risk-free security exists. The graphic representation of the new efficiency frontier can easily be found from its version with only risky securities. We will subsequently derive it analytically.

For one dollar to invest,  $x_*$  dollars can be put into the risk-free security (standard error=0, mean= $R_*$ ), and  $(1 - x_*) = \sum_{k=1}^{K} x_k$  dollars into some

risky portfolio with standard error  $\hat{\sigma}$  and mean  $\hat{m}$ . This operation yields a portfolio with standard error and mean given by:

$$\sigma = |1 - x_*|\hat{\sigma}, \quad m = x_*R_* + (1 - x_*)\hat{m}.$$

Consider first an investment  $x_*$  in the risk-free security that is less than one, implying that the value of the risky part of the portfolio is positive. Graphically, for a given  $(\hat{\sigma}, \hat{m})$ , the point  $(\sigma, m)$  describes the ray originating at the point representing the risk-free security  $(0, R_*)$  and passing through  $(\hat{\sigma}, \hat{m})$ . Now, if we consider all possible risky portfolios,  $(\hat{\sigma}, \hat{m})$  varies within a zone delimited by the hyperbola in Figure 7 and the set of rays describes the cone resting on the efficiency frontier.

Consider now an investment  $x_*$  greater than 1: the cone supported by the branch of the hyperbola that is symmetric with respect to the *m*-axis is obtained.

Mean-variance efficient combinations are located on the upper frontier of this cone: it is the ray that is tangent to the hyperbola which originates at the point representing the risk-free security. The point of tangency corresponds to a portfolio made only of risky securities  $x^*$ . This shows that all efficient portfolios are linear combinations of  $x^*$  and the risk-free security:

#### Theorem 4.1 The two-fund theorem

Assume there is a riskless security. For given expected returns and covariance matrix, all mean-variance efficient portfolios can be made up from two pooled investment funds: the risk-free security and a single risky fund.

These results can easily be derived analytically. Let  $(x_*, x)$  be a portfolio composition, where x is the risky component and  $x_*$  the risk-free component. Efficient compositions are solutions to

$$\min x' \Gamma x \text{ s.t.} \begin{cases} x_* R_* + E x' \tilde{R} \geq m \\ x_* + \mathbb{1}'_K x = 1, \end{cases}$$

and the first-order conditions are written

$$\Gamma x = \lambda E \tilde{R} + \mu \mathbf{1}_K \quad \text{and} \quad 0 = \lambda R_* + \mu,$$
(4.2)

whence, eliminating  $\mu$ 

$$\Gamma x = \lambda (ER - R_* \mathbf{1}_K). \tag{4.3}$$

Assume first that  $\lambda$  is non-nil and that the expected-return constraint is binding. We immediately see that the risky component of the portfolio is independent of the target expectation m and proportional to

$$x^* = \alpha \Gamma^{-1} (E\tilde{R} - R_* \mathbf{1}_K), \qquad (4.4)$$

where  $\alpha$  is set so as to normalize  $x^*$ . The expected return m determines the amount  $x_*$  invested in the risk-free security. Thus, the optimal portfolio consists of  $x_*$  invested in the risk-free security and  $(1 - x_*)$  in portfolio  $x^*$ .

If  $\lambda$  is nil, then so are  $\mu$  and x. We invest everything in the risk-free security, which only works if the target expectation is less than  $R_*$ —in this case the variance is minimum, equal to zero.

#### Remark

In practice, the return of the riskless security,  $R_*$ , is less than b/d. This is the case pictured in Figure 7. All efficient portfolios contain (positive) investments in risky securities, and the amount invested in them increases with the portfolio expected return. It is theoretically possible for  $R_*$  to be greater than b/d. Then,  $x_*$  is greater than 1 for all efficient portfolios, and investors take short positions on risky securities. Here again, the greater the absolute value of the position in risky securities, the greater the expected return.

# 2 Portfolio choice under the von Neumann Morgenstern criterion

We revert to the framework introduced in Chapter 3, assuming that the investor's tastes are represented by a von Neumann Morgenstern utility function,  $Ev(\tilde{c})$ . The function v is increasing, concave, and twice differentiable everywhere. Since a portfolio  $(z_*, z)$  yields a stochastic payoff in period 1

$$\tilde{c} = z_* + \sum_{k=1}^{K} z_k \tilde{a}_k = z_* + z' \tilde{a},$$

the portfolio choice is described by the following program:

An investor chooses a portfolio  $(z_*, z)$  maximizing

$$Ev(\tilde{\omega} + z_* + z'\tilde{a})$$

subject to the budget constraint:

$$p_*z_* + p'z = \omega_0.$$

It is convenient to define the "indirect" utility function,  $V(z_*, z)$ , derived from the "primal" utility function and containing the decision variables:

$$V(z_*, z) \equiv Ev(\tilde{\omega} + z_* + z'\tilde{a}).$$

The investor's program is then written:

$$\max_{z_*,z} V(z_*,z)$$
, s.t.  $p_*z_* + p'z = \omega_0$ . (4.5)

In contrast to the mean-variance framework, this program is not in general homogeneous with respect to wealth. By construction, the function V inherits the properties of being increasing and concave from the function v(note, however, that the domain of portfolios over which V is defined may depend upon the matrix of incomes  $\tilde{a}$ ). Thus, this problem has the traditional structure of a consumer's utility maximization subject to a budget constraint. There is one important difference: the domain of maximization may be unbounded because sales are allowed without restriction. Therefore the existence of a solution is not guaranteed. When security prices offer opportunities for arbitrage, the investor benefits by taking short positions on an expensive security to finance a purchase of an inexpensive one. Without limits on trades, no optimal solution exists, as shown in Theorem 3.2.

#### Remark

Working directly with future contingent income, rather than using the intermediary of the portfolio, is possible. It is convenient if markets are complete and without arbitrage opportunities. The vectors of security payoffs generate the entire space of contingent incomes (E = K + 1). As seen in Chapter 2, working directly with disposable income in each state of nature is equivalent to dealing with contingent goods whose prices are equal to the state prices. Thus, we end up with a formulation that is identical to that of the consumer in traditional microeconomic theory. Also, dropping possible redundant securities, portfolios are in a one-to-one correspondence with future incomes. When markets are incomplete, which is a possibility we do not want to preclude, the set of attainable incomes is constrained. Using future incomes as variables forces us to account for E - K - 1 additional constraints, and it is just as easy to work with portfolios.

The investor's program (4.5) consists of maximizing a concave function on a convex set. Thus, the first-order conditions characterize the solution. Letting  $\lambda$  be the multiplier associated with the budget constraint at time 0, they are written

$$\begin{cases} Ev'(\tilde{c}) = \lambda p_* \\ Ev'(\tilde{c})\tilde{a}_k = \lambda p_k, \quad k = 1, \dots, K. \end{cases}$$

$$(4.6)$$

As expected, these conditions can be satisfied only if there are no opportunities for arbitrage. Indeed they yield the investor's 'implicit' state prices.<sup>4</sup> To see this, writing expectation as a sum, (4.6) gives:

$$\sum_{e=1}^{E} \pi(e) v'[c(e)] a_k(e) = \lambda p_k, \quad k = *, 1, \dots, K.$$

 $<sup>^{4}</sup>$ These state prices are those for which the investor would chose the same contingent income profile if all contingent markets were to exist.

This shows, comparing with Theorem 2.1, If an optimal portfolio exists, there are no opportunities for arbitrage, and the positive vector q defined by

$$q(e) = \pi(e)v'(c_e)\lambda$$

is a vector of state prices. It satisfies the equalities

$$\sum_{e=1}^{E} q(e)a_k(e) = p_k, \quad k = *, 1, \dots, K.$$

If markets are complete, this last system of equations is of full rank and state prices are unique. All investors' vectors of marginal utilities for the contingent goods are thus proportional to each other.

To find a solution, note that completing the first order conditions with the budget constraint, a system with K + 2 equations for solving K + 2variables  $(z_*, z, \lambda)$  is obtained. Eliminating the multiplier  $\lambda$  and using the equality  $p_* = 1/(1+r)$ , an optimal portfolio is thus characterized by

$$\begin{cases} Ev'(\tilde{c})[\tilde{a}_k - (1+r)p_k] = 0, \quad k = *, 1, \dots, K, \\ p_* z_* + p' z = \omega_0. \end{cases}$$
(4.7)

# 3 Finance paradigms: quadratic and CARA-normal

In two specifications widely used in finance, quasi-explicit expressions can be derived for savings and the demands of securities. The determinants of the demand for financial securities can be easily interpreted. Furthermore, a link is established with the mean-variance criterion.

In the first specification, a quadratic utility function is assumed:

$$v(c) = c - \frac{\alpha}{2}c^2,$$

where  $\alpha$  is sufficiently small so that the function is increasing in c in the relevant domain.

The identity  $E\tilde{c}^2 = (E\tilde{c})^2 + \operatorname{var}(\tilde{c})$  allows to write

$$Ev(\tilde{c}) = E(\tilde{c}) - \frac{\alpha}{2}(E\tilde{c})^2 - \frac{\alpha}{2}\operatorname{var}(\tilde{c}),$$

or

$$Ev(\tilde{c}) = v(E\tilde{c}) - \frac{\alpha}{2} \operatorname{var}(\tilde{c}).$$
(4.8)

In the second specification, referred to, for short, as CARA-normal, utility exhibits constant absolute risk aversion (CARA)

$$v(c) = -\exp(-\rho c),$$

with  $\rho$  positive, and the distribution of payoffs is normal. Using the standard formula for the expectation of a log-normal variable, the expected utility of income during the second period is written

$$Ev(\tilde{c}) = -\exp\left\{-\rho\left[E\tilde{c} - \frac{\rho}{2}\operatorname{var}(\tilde{c})\right]\right\} = v\left[E\tilde{c} - \frac{\rho}{2}\operatorname{var}(\tilde{c})\right].$$
(4.9)

In both cases, the agent's utility is increasing in the expectation of future income  $E\tilde{c}$  and decreasing in its variance  $var(\tilde{c})$ . It is independent of the other moments of future income.

The expectation and variance of income associated with the purchase of portfolio  $(z_*, z)$  can be easily computed, using the expression  $\tilde{c} = \tilde{\omega} + z_* + z'\tilde{a}$ . The linearity of the expectation and the bilinearity of the covariance give:

$$E\tilde{c} = E\tilde{\omega} + z_* + z'E\tilde{a}$$
  
var( $\tilde{c}$ ) =  $z'$  var( $\tilde{a}$ )  $z + 2z'$ cov( $\tilde{a}, \tilde{\omega}$ ) + var( $\tilde{\omega}$ ).

In both the quadratic and CARA-normal cases, the optimal portfolio is the solution of a program

$$\max V(z_*, z) = f[E\tilde{c}, \operatorname{var}(\tilde{c})] \text{ s.t. } p_* z_* + p' z = \omega_0,$$

for some function f, where  $E\tilde{c}$  and  $var(\tilde{c})$  are given by their expressions in terms of  $(z_*, z)$ .

The important point to note here is that total income, not only financial income, matters. In the presence of risky non financial income, the optimal portfolio is generally not mean-variance efficient because the variance of total income,  $\tilde{\omega} + z_* + z'\tilde{a}$ , differs from that of financial income,  $z_* + z'\tilde{a}$ . However,

when the non financial income is risk free (or uncorrelated with the securities payoffs), the quadratic and CARA-normal specifications both lead the investor to choose a mean-variance efficient portfolio.

# 3.1 Hedging portfolios

#### **Definition 3**

A portfolio that minimizes the variance of total income in period 1 is called a hedging portfolio.

The variance only depends on the risky securities in the portfolio, and the hedging portfolio  $z^h$  minimizes:

$$\operatorname{var}(\tilde{\omega} + z'\tilde{a}) = z' \operatorname{var}(\tilde{a}) z + 2z' \operatorname{cov}(\tilde{a}, \tilde{\omega}) + \operatorname{var}(\tilde{\omega}).$$

This portfolio is the one that would be chosen by an individual who is infinitely averse to variance in income ( $\alpha$  or  $\rho$  equal to  $+\infty$  in the quadratic and CARA-normal models, respectively). Thus, its value is

$$z^{h} = -\operatorname{var}(\tilde{a})^{-1}\operatorname{cov}(\tilde{a},\tilde{\omega}).$$
(4.10)

This result has a geometric interpretation. Note that we are only interested in the variance of incomes. Thus their expectations can be subtracted so as to work exclusively on centered incomes, that is in the (E - 1)dimensional space of variables with zero expectation. In this space, the covariance is a scalar product with associated norm the square root of the variance, that is the standard error. The centered incomes that are attainable with these portfolios define a subset spanned by the centered risky security payoffs,  $\tilde{a}_k - (E\tilde{a}_k)$ , in a K-dimensional subspace.<sup>5</sup> Hedging con-

 $<sup>^{5}</sup>$ Owing to the assumption of no redundancy, the vector made of one, 1<sub>\*</sub>, is independent

sists of choosing an income in this subspace that reduces the variance of total income to the greatest possible extent. This amounts to a projection. By definition, the projection of a vector onto a subspace is the closest vector of that subspace. Income is thus decomposed into the sum of its projection and a vector that is orthogonal to the subspace. Using the variance as the square of the distance between random variables, we can write for any centered  $\tilde{x}$ :

$$\tilde{x} = \operatorname{proj}_{\tilde{a}} \tilde{x} + \tilde{x}_{\perp}$$
 with  $\operatorname{proj}_{\tilde{a}} \tilde{x} = \operatorname{cov}(\tilde{x}, \tilde{a}) \operatorname{var}(\tilde{a})^{-1} (\tilde{a} - E\tilde{a})$ 

where:

- $\operatorname{proj}_{\tilde{a}}\tilde{x}$  is a linear combination of the payoffs  $\tilde{a}_k E\tilde{a}_k$  and can thus be attained by a portfolio.
- $\tilde{x}_{\perp}$  is in a subspace that is orthogonal to security payoffs (by construction, its covariance with  $\tilde{a}$  is equal to 0). It is not correlated with security payoffs and consequently cannot be insured by the market.

Applying this result to non financial income  $\tilde{\omega} - E\tilde{\omega}$ , its projection onto security payoffs is from (4.10) exactly the opposite of the payoffs of the hedging portfolio, that is  $-z^{h'}(\tilde{a} - E\tilde{a})$ . Therefore, after hedging, final income is uncorrelated with security payoffs.<sup>6</sup> If markets are complete, K = E - 1, the investor can completely insure herself against non financial risks: By selling the portfolio that duplicates non financial income, the residual variance is nil. When markets are incomplete, some of the non financial risks typically remain uninsurable on the markets.

$$\operatorname{cov}[\tilde{a}, \tilde{\omega} - E\tilde{\omega} - z'_h(\tilde{a} - E\tilde{a})] = \operatorname{cov}(\tilde{a}, \tilde{\omega}) - \operatorname{var}(\tilde{a})z_h = 0.$$

of the vectors  $\tilde{a}_k$ , k = 1, ..., K. Thus, by centering, the dimension of the spanned space is not reduced, and remains equal to K.

<sup>&</sup>lt;sup>6</sup>This can be confirmed by calculating:

#### **3.2** The demand for risky securities

The calculation of the demand differs slightly in the quadratic and CARAnormal cases.

Let us begin with the quadratic case. Recall the first-order conditions (4.7) for the K risky securities:

$$Ev'(\tilde{c})[\tilde{a}_k - (1+r)p_k] = 0,$$

which can be also written as:

$$Ev'(\tilde{c})E\tilde{a}_k + \operatorname{cov}[v'(\tilde{c}), \tilde{a}_k] = (1+r)p_k Ev'(\tilde{c}).$$

Since  $v'(\tilde{c}) = 1 - \alpha \tilde{c}$ , we have

$$Ev'(\tilde{c}) = v'(E\tilde{c}) \text{ and } \operatorname{cov}[v'(\tilde{c}), \tilde{a}_k] = -\alpha \operatorname{cov}(\tilde{\omega} + z'\tilde{a}, \tilde{a}_k).$$

Stacking up these equations gives

$$v'(E\tilde{c})[E\tilde{a} - (1+r)p] - \alpha[\operatorname{var}(\tilde{a})z + \operatorname{cov}(\tilde{a},\tilde{\omega})] = 0.$$

Factoring out  $v'(E\tilde{c})$ , yields

$$z = \operatorname{var}(\tilde{a})^{-1} \left\{ -\operatorname{cov}(\tilde{a}, \tilde{\omega}) + \frac{v'(E\tilde{c})}{\alpha} \left[ E\tilde{a} - (1+r)p \right] \right\}$$

or, using the expression for risk tolerance  $T(c) = v'(c)/\alpha$  (see Chapter 3, Section 3.8)

$$z = \operatorname{var}(\tilde{a})^{-1} \{ -\operatorname{cov}(\tilde{a}, \tilde{\omega}) + T(E\tilde{c}) [E\tilde{a} - (1+r)p] \}.$$
(4.11)

The CARA-normal case yields the same equation, but the calculations are simpler, starting directly from the investor's problem. The investor maximizes  $v \{E\tilde{c} - \rho/2[var(\tilde{c})]\}$  subject to the budget constraint. Substituting  $z_*$  from the budget constraint as a function of z, the first-order condition in z yields Equation (4.11), using the fact that risk tolerance is equal to  $1/\rho$  for CARA utility functions.

Thus, the investor's demand for securities appears as the sum of two portfolios:

- The first is the hedging portfolio,
- The second is proportional to var(ã)<sup>-1</sup> [Eã (1 + r)p]. It is called speculative because it coincides with the demand of an investor with no risky non financial income to hedge (ῶ constant). By intervening in the market, he is taking some risk. When security payoffs are not correlated and have the same variance, the speculative portfolio is proportional to Eã (1 + r)p, the vector of the expected security payoffs in excess of discounted prices (the discounting is necessary because prices are paid at time 0 and payoff received in the next period.) The investor buys a security if the expected payoff is larger that the discounted price and sells it if it is lower.

Thus, in the two specifications just examined, the portfolio choice responds to two goals, risk reduction and speculation. Clearly, the relative weights assigned to each of these elements depend on the amount of initial risk to hedge and the opportunities reflected by the securities, but also on risk tolerance: the speculative part increases with the investor's risk tolerance.

The difference with the mean-variance case merits emphasis. The twofund theorem is only applicable to speculative demand: Speculators with the same expectations choose the same portfolios, and they behave according to the mean-variance model. Instead, the composition of the hedging portfolio depends on individual non financial incomes and usually vary from one investor to the other. Overall, in this model any difference in portfolio composition must be attributable to variations in hedging requirements or to heterogeneous beliefs on the securities payoffs.

#### Remarks

1. The expression (4.11) for the optimal portfolio is not entirely explicit in the quadratic case, since risk tolerance depends on total expected income, and thus on the investment on the risk-free security and ultimately on the interest rate.

2. When we move beyond these two simple specifications, the first-order conditions in (4.6) are no longer linear in z and the demand for securities simultaneously reflects hedging needs and a desire for profitable investments—though it is usually not feasible to distinguish between these two elements.

#### **Bibliographical Note**

Mean-variance analysis, founded by Markowitz' (1952), is very popular in finance. The presentation here is similar to that in Ingersoll (1987). The volume edited by Diamond and Rothschild (1978) contains applications and critical insights on the von Neumann Morgenstern theory.

Diamond, P. et M. Rothschild (1978) : *Uncertainty in Economics*, Academic Press, Boston.

Ingersoll, J.E. Jr. (1987) : *Theory of Financial Decision Making*, Rowman et Littlefield, New York.

Markowitz, H. (1952) : "Portfolio Selection", Journal of Finance, 7, 77-91.

#### Exercises

#### **Exercise 7**: The mean-variance criterion

Consider a stock market with two risky securities (k = 1,2). Their returns (per dollar invested) are denoted by  $\tilde{R}_k$ , k = 1,2, with mathematical expectation and standard deviation  $m_k$  and  $\sigma_k$ , respectively. Let  $\rho$  be the correlation coefficient between  $\tilde{R}_1$  and  $\tilde{R}_2$ , i.e.  $\rho = \operatorname{cov}(\tilde{R}_1, \tilde{R}_2)/(\sigma_1\sigma_2)$ . There is also a riskless security (k = \*) that yields  $R_* = (1 + r)$  per dollar invested.

Let  $\hat{R}$  be the return of a portfolio of composition x, i.e.:

$$\ddot{R} = x_* R_* + x_1 \ddot{R}_1 + x_2 \ddot{R}_2,$$

in which  $x_k$  represents the fraction invested in k (k = \*, 1, 2):

$$x_* + x_1 + x_2 = 1.$$

There is no condition on the sign of the  $x_k$ , implying that short selling of any security is allowed. Finally, denote the mathematical expectation and the standard deviation of  $\tilde{R}$  by m and  $\sigma$ , respectively.

- 1. Show that all portfolios that consist exclusively of risky securities define a curve in the plane  $(\sigma, m)$ . Represent this curve by assuming  $m_1 = 2, m_2 = 1, \sigma_1 = 2, \sigma_2 = 1$ , in the cases  $\rho = 1, \rho = 1/2, \rho = 0$  and  $\rho = -1$ . Identify the part of the curve that can be reached when short sales are not allowed. Comment.
- 2. Assume that  $\rho = 1/2$  and  $m_* = 1/2$ . Find the equation of the efficiency frontier. What is the composition of the mutual fund of risky securities chosen by any investor? Comment.
- 3. Characterize the demand of securities of an investor with preferences represented by a utility function  $U(m, \sigma) = m \alpha \sigma^2, \alpha > 0$ .

#### **Exercise 8**: Maximizing utility and the mean-variance criterion

Under some circumstances the quadratic utility function leads to the choice of mean-variance efficient portfolios. However, it has two drawbacks: First, the possible values for income must be restricted to the domain on which the function is increasing (recall that a concave quadratic function is surely decreasing for sufficiently high values of its argument). But most of all, absolute risk aversion increases with the level of wealth in the case of the quadratic function, which violates current observations.

An alternative rationale for mean-variance analysis makes use of assumptions that are not on the form of the utility function, but rather on the distribution of portfolio returns.

## I. The normal case

An investor invests wealth  $\omega_0$  to acquire a portfolio z comprised of K securities. With the usual notations,  $p'z = \omega_0$ , and the portfolio yields a random financial income of  $z'\tilde{a} = \sum_{k=1}^{K} z_k \tilde{a}_k$  at time 1.

The investors' preferences over financial incomes are represented by a von Neumann Morgenstern utility index  $v : \mathbb{R} \to \mathbb{R}$ —a strictly increasing, strictly concave, and twice continuously differentiable function.

Let g denote the probability density function of the standard normal distribution:

$$g(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

To ensure that the integrals are well defined, we assume that there exists a number M such that, for all m and  $\sigma$ , the functions f(x) = v(x), v'(x) or v''(x) satisfy:

$$\lim_{x \to \pm \infty} |x^2 f(m + \sigma x)g(x)| \le M.$$

1. Check that the choice of a portfolio only depends on the mean and variance of its payoffs. Denoting by  $(m, \sigma)$  the couple (mean, standard error) associated with the portfolio z, define:

$$V(m,\sigma) = Ev(z'\tilde{a}).$$

Show that:

$$\frac{\partial V}{\partial m} = Ev'(z'\tilde{a}), \text{ and } \frac{\partial V}{\partial \sigma} = \sigma Ev''(z'\tilde{a}).$$

Demonstrate that the function V is concave in  $(m, \sigma)$ . What is the form of the indifference curves of V in the  $(\sigma, m)$  plane?

2. Compute the optimal portfolio for a utility function with constant absolute risk aversion  $\rho > 0$ :

$$v(x) = -\exp(-\rho x).$$

Assume that there exists a risk free security . How do investor's choices vary with their risk aversions?

#### II. Non-normal distributions

The preceding analysis relies on two properties: (1) Normal distributions are characterized by two parameters, and (2) Any linear combination of normal distributions is a normal distribution.

The mean-variance analysis extends to families of random variables whose distributions depend on two parameters and that are stable by linear combination. There are many such families besides the multivariate normal distribution: They are called elliptical distributions. We examine here an example. Let  $(\tilde{\alpha}, \tilde{\beta})$  be a couple of real random variables with density function  $h(\alpha, \beta)$  where  $\tilde{\beta}$  is strictly positive with probability 1.

- 1. Consider the family of random variables  $\tilde{a}$ , whose conditional distribution given  $(\alpha, \beta)$  is a normal distribution with mean  $\alpha M$  and variance  $\beta \Sigma^2$ , for a couple  $(M, \Sigma)$  of  $\mathbb{R} \times \mathbb{R}_+$ .
  - (a) Demonstrate that the family is stable by linear combination.
  - (b) Compute the unconditional expectation and variance of  $\tilde{a}$ . Show that, if  $E\tilde{\alpha} \neq 0$ , the unconditional expectation and variance characterize the distribution within the family studied. Does this property extend to the multivariate case?
- 2. Consider an investor with constant absolute risk aversion  $\rho$ . In the specific case in which  $\alpha = 1$  and  $\tilde{\beta}$  has an exponential distribution with parameter  $\mu$ ,

$$h(\beta) = \mu \exp(-\mu\beta),$$

compute the indirect utility function  $V(m, \sigma)$  where  $(m, \sigma)$  represent the unconditional mean and standard error of a portfolio payoff. Specify the domain over which it is defined. Derive the portfolio demand. Compare it with the normal case above.

**Exercise 9**: Speculative and hedging demands

Let a security be priced p at time t = 0 with an expected payoff at t = 1 (future price + dividend) denoted by  $\tilde{a}$ . The risk free interest rate between t = 0 and t = 1 is r. We consider an investor whose preferences are represented by a strictly concave von Neumann Morgenstern criterion v on income at date 1.

1. The investor's initial wealth is composed of  $w_0$  units of money. If she purchases z shares of the security at t = 1, her expected wealth at t = 1 is

$$\tilde{w} = (1+r)w_0 + (\tilde{a} - (1+r)p)z$$

Write the first-order conditions characterizing the demand of security. Show that the demand is

- $-\operatorname{nil} \operatorname{if} (1+r)p = E\tilde{a}$
- positive if  $(1+r)p < E\tilde{a}$
- negative if  $(1+r)p > E\tilde{a}$ .
- 2. Now assume that the investor initially has  $w_0$  units of money and  $z_0$  of the security. Using 1), decompose her demand into a speculative and a hedging component.