6

Stochastic methods in economics and finance

Vassili N. Kolokoltsov

Abstract

Economic behavior and market evolution present notoriously difficult complex systems, where physical interacting particles become purpose-pursuing interacting agents, thus providing a kind of a bridge between physics and social sciences.

We systematically develop the mathematical content of the basic theory of financial economics that can be presented rigorously using elementary probability and calculus, that is, the notions of discrete and absolutely continuous random variables, their expectation, notions of independence and of the law of large numbers, basic integration – differentiation, ordinary differential equations and (only occasionally) the method of Lagrange multipliers. We do not assume any knowledge of finance, apart from an elementary understanding of the idea of compound interest, which can be of two types: (i) simple compounding with rate r and a fixed period of time means your capital in this period is multiplied by $(1+r)$; (ii) continuous compounding with rate r means your capital in a period of time of length t is multiplied by e^{rt}.

This chapter is based on several lecture courses for statistics and mathematics students at the University of Warwick and on invited mini-courses presented by the author at various other places. Sections 6.2 and 6.3 are developed from the author's booklet [9]. The chapter is written in a rather concise (but comprehensive) style in attempt to pin down as clear as possible the mathematical relations that govern the laws of financial economics. Numerous heavy volumes are devoted to the detail

Complexity Science: The Warwick Master's Course, eds. Robin Ball, Vassili Kolokoltsov and Robert S. MacKay. Published by Cambridge University Press.

discussion of the economic content of these mathematical relations, see e.g. [5], [6], [8], [15], [17].

6.1 Utility theory

Here we sketch the basics of utility theory presenting the main definitions, but omitting the proofs of key theorems (which do not contribute much to their economic content). The full story can be found in [1].

6.1.1 Order, utility and expected utility

We start with an abstract definition forming the basis for the theory.

Let M be a subset of a Euclidean space. A *binary relation* R on M is defined abstractly just as a subset of the product $M \times M$, i.e. a subset of the set of pairs (x, y), $x, y \in M$. One writes xRy, if the pair (x, y) belongs to this subset. With some ambiguity, we shall normally use the standard symbol $\leq$ for an abstract relation R.

The relation $\leq$ is called *transitive*, if $x \leq y$ and $y \leq z$ imply $x \leq z$, *reflexive*, if $x \leq x$ for all x, *symmetric* if $x \leq y \iff y \leq x$, *anti-symmetric*, if $x \leq y$ and $y \leq x$ imply $x = y$, *complete* if either $x \leq y$ or $y \leq x$ hold for any pair x, y, *continuous*, if the sets $\{x : x \geq y\}$ and $\{x : x \leq y\}$ are closed.

An *equivalence relation* is a symmetric, reflexive and transitive relation.

A *pre-order* is a reflexive and transitive relation.

If $\leq$ is a pre-order, one defines associated relations $x < y$ meaning that $x \leq y$, but not $y \geq x$. Also $x \geq y$ means $y \leq x$.

A *partial order* is a reflexive, transitive and anti-symmetric relation. A complete partial order is called a *complete order*.

Main examples of partial orders in $\mathbf{R}^d$ are the *Pareto order*, where $x = (x_1, \ldots, x_d) \leq (y_1, \ldots, y_d)$ means $x_j \leq y_j$ for all j, and *lexicographic order*, where $x = (x_1, \ldots, x_d) \leq (y_1, \ldots, y_d)$ means that either $x_1 < y_1$, or $x_1 = y_1$ and $x_2 < y_2$, or $x_1 = y_1$, $x_2 = y_2$ and $x_3 < y_3$, etc. Clearly Pareto order is not complete and lexicographic order is not continuous.

Utility functions were introduced in an attempt to measure preferences (pre-orders) quantitatively. One can say that a *utility function* is a measure of satisfaction. For any function U on a set M one can define the relation $x \leq y$ that holds if and only if $U(x) \leq U(y)$ (the latter relation is the usual order on real numbers). Function U is then called the *utility function* corresponding to (or defining) this pre-order. For any U this relation is a complete pre-order. If U is continuous, this pre-order is continuous. Remarkably enough, the inverse statement is also true, as the following theorem claims.

Theorem 6.1.1 Main theorem on utility functions. *If a pre-order on* $\mathbf{R}^d$ *is complete and continuous, it can be specified via a continuous (utility) function.*

In economics it is vital to compare uncertain outcomes, leading to the necessity to define pre-orders on random variables. The simplest random variables are those with a finite range, which can be identified with what is usually referred to as simple lotteries. Let Y be a subset of a Euclidean space. A *simple lottery* on Y is a finite convex combination of Dirac's point measures:

$$p = \sum_{y \in \mathrm{supp(p)}} p(y)\delta_y, \qquad \sum_{y \in \mathrm{supp(p)}} p(y) = 1,$$

so that the measure p is identified with a function $p : Y \to [0, 1]$ with a finite support supp(p) (economists' notation $\sum_{y \in \mathrm{supp(p)}} p(y)1_y$). If supp(p) is one point, p is called a *degenerate lottery*. Countable mixtures of Dirac's point measures, called *discrete lotteries*, are also sometimes used in economics, but much more rarely, and we shall not touch them here.

Let us denote by $\Delta(Y)$ the set of all simple lotteries.

Any utility function v on Y induces a utility on $\Delta(Y)$ called *expected utility*:

$$\mathbf{E}v(p) = \sum_{y \in \mathrm{supp(p)}} p(y)v(y) = \int_Y v(y)p(dy).$$

Clearly, a relation on $\Delta(Y)$ specified by expected utility enjoys the following properties:

(O) ordering: $\leq$ is a complete pre-order on $\Delta(Y)$;
(I) independence: $\lambda > \mu \implies \alpha\lambda + (1-\alpha)\nu > \alpha\mu + (1-\alpha)\nu$ for all $\alpha \in (0,1)$, $\lambda, \mu, \nu \in \Delta(Y)$;
(C) continuity: if $\lambda, \mu, \nu \in \Delta(Y)$ and $\lambda > \mu > \nu$, then $\exists \alpha, \beta \in (0,1)$:

$$\alpha\lambda + (1-\alpha)\nu > \mu > \beta\lambda + (1-\beta)\nu;$$

Theorem 6.1.2 Main theorem on expected utility. *If a relation* $\leq$ *on* $\Delta(Y)$ *satisfies (O), (I), (C), then there exists a function* v *on* Y*, unique up to a linear equivalence (also called ordinal equivalence in this context), such that* $\leq$ *is generated by the expected utility* $\mathbf{E}v$.

6.1.2 Utility on monetary outcomes

Let a real function u on $\mathbf{R}$ or $\mathbf{R}_+$ be twice differentiable.

One says that u satisfies the *principle of non-satiation*, if $u' > 0$; u is *risk averse* (resp. *risk seeking* or *risk indifferent*) if $u'' < 0$ (resp. $u'' > 0$ or $u'' = 0$). The derivative u' designates the rate of growth of utility with the increase of monetary outcomes. Risk aversion means that when your wealth increases, your satisfaction from a given incremental increase decreases (e.g. when you have \$1 or \$1000, your interest in another \$1 is quite different).

Another interpretation: let u be risk averse (that is increasing and concave), $p \in (0,1)$, and P the lottery winning a or b with probabilities p or $1-p$ resp., with expectation $\mathbf{E}(P) = pa + (1-p)b$. Then

$$\mathbf{E}u(P) = pu(a) + (1-p)u(b) \le u(pa + (1-p)b) = u(\mathbf{E}(P)), \quad (6.1)$$

that is, the expected utility of a lottery with the expectation $c = \mathbf{E}(P)$ is less then the utility of the sure $\mathbf{E}(P)$.

Certainty equivalent $s(P)$ of a lottery P is defined via $u(s(P)) = \mathbf{E}u(P)$. By (6.1), $s(P) \le \mathbf{E}(P)$. See Fig. 6.1.

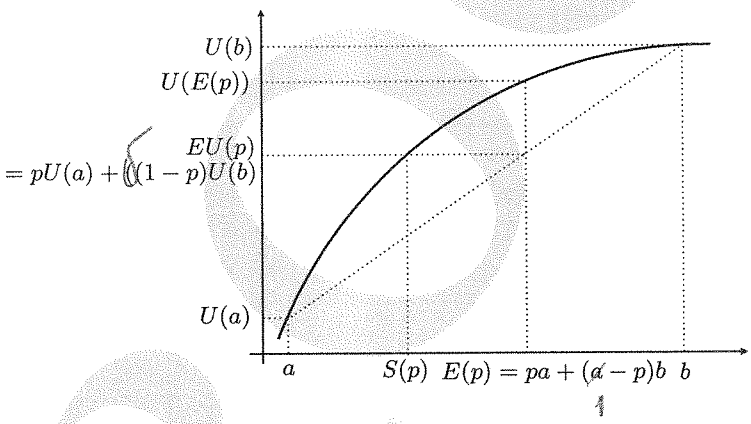

Figure 6.1 Certainty equivalent.

Utility functions can be quantitatively compared by their *absolute* and *relative risk aversions*, *ARA* and *RRA* respectively, which are defined as

$$ARA(u) = -\frac{u''(w)}{u'(w)}, \quad RRA(u) = -w\frac{u''(w)}{u'(w)}.$$

For a given ARA $\rho(x)$, there exists a unique utility with $u(0) = u'(0) =$

0, namely

$$u(x) = \int_0^x \exp\{-\int_0^y \rho(y)dy\}\, dx. \tag{6.2}$$

Basic examples of utility functions are the following. A utility function of form $u(x) = a + bx + cx^2$ is called quadratic (of course). A utility function is called *CARA* (constant ARA) if its ARA is a constant. For a CARA utility, the acceptance of a gamble g does not depend on an initial capital, that is the validity of the inequality $\mathbf{E}u(x+g) > u(x)$ does not depend on the constant x. A utility is called *CRRA* (constant RRA) if its RRA is a constant. A utility is called *HARA* (hyperbolic ARA), if its ARA is inverse to linear, that is of the form $1/(ax+b)$ with constants a, b.

6.1.3 Dominance

Expected utilities allows one to compare random variables. But utility functions are difficult to measure. A question arises whether one can characterize intrinsically the order arising from all utilities satisfying non-satiation principle and/or risk aversion.

For two random variables A and B with distribution functions F_A and F_B, A is said to *stochastically dominate B in the first order* non-strictly (denote $A \geq_1 B$), if for all x

$$F_A(x) \leq F_B(x) \Longleftrightarrow \mathbf{P}(A > x) \geq \mathbf{P}(B > x),$$

and strictly (denote $A >_1 B$) if additionally B does not dominate A, that is there exists x such that $F_A(x) < F_B(x)$.

If $A \geq_1 B$, then there exist two random variables A', B' on one and the same probability space such that A' (resp. B') has the same distribution as A (resp. B) and $A' \geq B'$ point-wise. In fact, one can take $A' = F_A^{-1}$, $B' = F_B^{-1}$ (see Proposition 6.8.1 for the inverse function method employed here).

For two random variables A and B with range $[a, b]$ and distribution functions F_A and F_B, A *stochastically dominates B in the second order* (or in the sense of the *stop-loss stochastic order*) non-strictly (denote $A \geq_2 B$), if for all x

$$\Phi_A(x) = \int_a^x F_A(y)dy \leq \Phi_B(x) = \int_a^x F_B(y)dy,$$

and strictly (denote $A >_2 B$) if additionally B does not dominate A, that is there exists x such that $\Phi_A(x) < \Phi_B(x)$.

It is easy to see that the first- and second-order dominance specify partial orders on the set of random variables.

For a random variable A and a utility function u we shall write shortly $u(A)$ for $\mathbf{E}u(A)$.

Proposition 6.1.3 Let U^1 be the set of all utility functions u on $[a,b]$ with $u' > 0$. Then

$$A \geq_1 B \Longleftrightarrow u(A) \geq u(B) \quad \forall u \in U^1,$$

$$A >_1 B \Longleftrightarrow u(A) > u(B) \quad \forall u \in U^1.$$

$u(A) \geq u(B)\ \forall u \in U^1$ and $\exists u \in U^1 : u(A) > u(B)$

Proof From integration by parts

$$u(A) - u(B) = -\int_a^b u'(x)(F_A(x) - F_B(x))dx,$$

implying what was claimed. □

Proposition 6.1.4 Let U^2 be the set of all utility functions u on $[a,b]$ with $u' > 0$ and $u'' < 0$. Then

$$A \geq_2 B \Longleftrightarrow u(A) \geq u(B) \quad \forall u \in U^2,$$

$$A >_2 B \Longleftrightarrow u(A) > u(B) \quad \forall u \in U^2.$$

$u(A) \geq u(B)\ \forall u \in U^1$ and $\exists u \in U^2 : u(A) > u(B)$

Proof

$$u(A) - u(B) = -\int_a^b u'(x)(F_A(x) - F_B(x))dx$$

$$= -u'(b)(\Phi_A(b) - \Phi_B(b)) + \int_a^b u''(x)(\Phi_A(x) - \Phi_B(x))dx,$$

implying what was claimed. □

6.1.4 Utility on $\mathbf{R}_+^d$

Any combination of available goods is referred to as a *consumption bundle*. An *indifference curve* (or *surface*) comprises all bundles of equal utility. A *marginal rate of substitution* is the amount of one good that a consumer is prepared to swap for one extra unit of another good. They are given by slopes of indifferent surfaces (or curves obtained from the intersections of an indifference surface with a two-dimensional section of $\mathbf{R}_+^d$).

A function u on $\mathbf{R}^d_+$ is called *quasi-concave* (resp. *strictly quasi-concave*), if $\{w : u(w) \geq w\}$ is convex for each w (resp. strictly convex).

Two interpretations of quasi-concavity are worthy of mention. Namely, if u is quasi-concave and increasing (in all variables), then:

(a) mixtures of bundles I and II of equal utility are not worse than I or II (see Fig. 6.2);

(b) consumer preference exhibits *diminishing marginal rates of substitution*: if a person is happy to exchange a cherry for an apple, it would take more than one apple to persuade him/her to give another cherry (see Fig. 6.3).

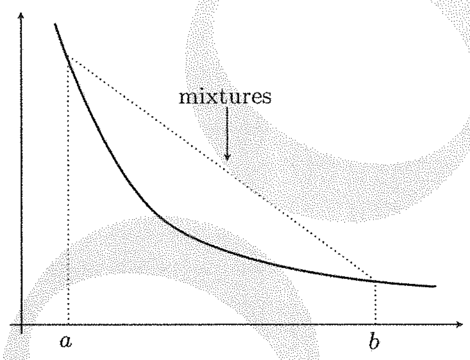

Figure 6.2 Mixtures of consumption bundles.

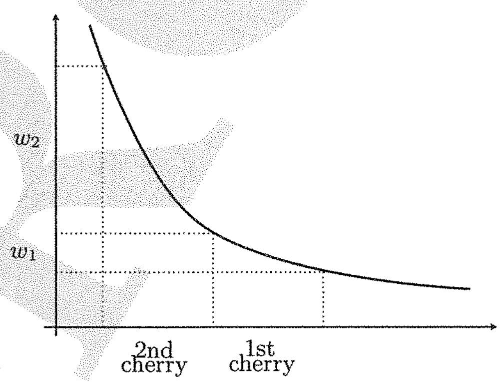

Figure 6.3 Diminishing marginal rate of substitution.

Consumer choice cost-function (or *expenditure function*) for a given set of prices $p \in \mathbf{R}_+^d$ is defined as

$$c(p, w) = \min_q \{(p, q) : u(q) \geq w\}. \tag{6.3}$$

If u is continuous, then c is a nondecreasing homogeneous and concave function of p. The dual function

$$u^*(q) = \max\{w : (p, q) \geq c(p, w) \forall p\}$$

allows us to reconstruct utility from the cost-function, due to the following fundamental result.

Theorem 6.1.5 Shephard's lemma and duality theorem. *If u is strictly quasi-concave, smooth and increasing (in all variables), then there exists a unique q^* minimizing the r.h.s. of* (6.3) *and*

$$q^* = \frac{\partial c}{\partial p}(p, u), \quad u^* = u.$$

6.2 Variance – spread – risk

6.2.1 Variance and correlation

Trying to describe a random variable X by some simple characteristics, one is naturally led to look for its location statistics and its spread, i.e. one looks for a point d, where the spread of the deviation $\mathbf{E}[(X - d)^2]$ is minimal and then assesses this spread.

As by the linearity of the mean

$$\mathbf{E}[(X - d)^2] = \mathbf{E}(X^2) - 2d\mathbf{E}(X) + d^2,$$

one easily finds that the minimum is attained when $d = \mathbf{E}(X)$, yielding a characterization of the expectation as the location statistics. The minimum itself

$$Var(X) = \mathbf{E}[(X - \mathbf{E}(X))^2] = \mathbf{E}(X^2) - [\mathbf{E}(X)]^2 \tag{6.4}$$

is called the *variance* of X and constitutes the second basic characteristics of a random variable describing its spread statistics. In applications, the variance often represents a natural measure of risk, as it specifies the average deviation from the expected level of gain, loss, win, etc. Variance is measured in square units, and the equivalent spread statistics measured in the same units as X is the *standard deviation* of X:

$$\sigma = \sigma_X = \sqrt{Var(X)}.$$

For instance, if $\mathbf{1}_A$ is the indicator function on a probability space, then $\mathbf{E}(\mathbf{1}_A) = \mathbf{P}(A)$ and $\mathbf{1}_A = (\mathbf{1}_A)^2$ so that

$$Var(\mathbf{1}_A) = \mathbf{E}(\mathbf{1}_A) - [\mathbf{E}(\mathbf{1}_A)]^2 = \mathbf{P}(A)[1 - \mathbf{P}(A)].$$

In particular, for a Bernoulli random variable X taking values 1 and 0 with probabilities p and $1 - p$ it implies

$$Var(X) = p(1 - p). \tag{6.5}$$

The simplest measure of the dependence of two random variables X, Y (forming a random vector, i.e with the joint probabilities specified) is supplied by their *covariance*:

$$Cov(X, Y) = \mathbf{E}[(X - \mathbf{E}(X))(Y - \mathbf{E}(Y))],$$

or equivalently by their *correlation coefficient*:

$$\rho = \rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

As easily follows from definition,

$$Cov(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y). \tag{6.6}$$

If $Cov(X, Y) = 0$, one says that the random variable X and Y are *uncorrelated*. If X and Y are independent, they are clearly uncorrelated. Generally speaking, the converse statement does not hold. However, it is possible to show (we will not go into detail here) that it does hold for Gaussian random variables, i.e. two jointly Gaussian random variables are uncorrelated if and only if they are independent. This implies, in particular, that the dependence structure of, say, two standard Gaussian random variables can be practically described by a singe parameter, their correlation. This property makes Gaussian random variable particularly handy when assessing dependence, as is revealed, say, by the method of Gaussian copulas, see Section 6.8.

The linearity of the expectation implies that for a collection of random variables $X_1, \ldots, X_k$ one has

$$Var(\sum_{i=1}^k X_i) = \sum_{i=1}^k Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j).$$

In particular, if random variables $X_1, \ldots, X_k$ are pairwise uncorrelated, for example if they are independent, the variance of their sum equals the sum of their variances:

$$Var(X_1 + \cdots + X_k) = Var(X_1) + \cdots + Var(X_k). \tag{6.7}$$

This linearity is very useful in practical calculations. For example, if X is a binomial (n,p) random variable, then

$$X = \sum_{i=1}^{n} X_i,$$

where X_i are independent Bernoulli trials, so that by (6.5) and by linearity one gets

$$Var(X) = np(1-p). \tag{6.8}$$

By the definition of the expectation for continuous random variables X described by a probability density function f, the variance can be calculated as

$$Var(X) = \int [x - \mathbf{E}(X)]^2 f_X(x)\, dx. \tag{6.9}$$

As an example, let us calculate the variance for a normal random variable. Recall first that a *normal or Gaussian* random variable is specified by the probability density functions of the form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \tag{6.10}$$

where $\mu \in \mathbf{R}$ and $\sigma > 0$ are two parameters. These random variables are usually denoted $N(\mu, \sigma^2)$. Random variables $N(0,1)$ are called *standard normal* and have the density

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}. \tag{6.11}$$

Its distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{y^2}{2}\right\} dy \tag{6.12}$$

can not be expressed in closed form, but is tabulated in standard statistical tables with great precision due to its importance. The properties of general $N(\mu, \sigma^2)$ are easily deduced from the standard normal, due to their simple linear connections. Namely, if X is $N(0,1)$, then

$$Y = \mu + \sigma X \tag{6.13}$$

is $N(\mu, \sigma^2)$. In fact,

$$F_Y(x) = \mathbf{P}(\mu + \sigma X \le x) = \mathbf{P}\left(X \le \frac{x-\mu}{\sigma}\right),$$

so that the distribution function of Y is

$$F_Y(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Differentiating with respect to x yields for the probability density of Y the expression (6.10) as was claimed.

Let X be $N(0,1)$. Then by (6.9) and (6.10) (taking into account that the expectation of X vanishes and using integration by parts)

$$Var(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp\left\{-\frac{1}{2}x^2\right\} dx$$

$$= -\frac{1}{\sqrt{2\pi}} x \exp\left\{-\frac{1}{2}x^2\right\}\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}x^2\right\} dx = 1,$$

as the second term is the integral of a probability density function. It follows from (6.13) that the variance of a $N(\mu, \sigma^2)$ normal random variable equals σ^2 and its expectation μ.

6.2.2 Volatility and correlation estimators

A standard problem in applications of probability is estimating the mean, variance and correlations of random variables on the basis of their observed realizations during a period of time. Such estimates are routinely performed, say, by traders, for assessing the volatility of the stock market or a particular common stock. In its simplest mathematical formulation the problem is to estimate the mean and variance of a random variable X, when a realization of a sequence $X_1, \ldots, X_n$ of independent random variables distributed like X was observed. It is of course natural to estimate the expectation μ of X by its empirical mean

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

This estimate is *unbiased* in the sense that $\mathbf{E}(\hat{\mu}_n) = \mu$ and *asymptotically effective* in the sense that $\hat{\mu}_n$ converges to μ by the law of large numbers. As the variance σ^2 is the expectation of $(X-\mu)^2$, the above reasoning suggests also the estimate for the variance in the form

$$\hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2,$$

which is again unbiased and asymptotically effective. The problem with this estimate lies in utilizing the unknown μ. To remedy this shortcoming, it is natural to plug in the above given estimate instead of μ leading to the estimate

$$\hat{\sigma}_2^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \hat{\mu}_n^2.$$

But here a surprise is awaiting. This estimate is no longer unbiased. In fact,

$$\mathbf{E}(\hat{\mu}_n^2) = \frac{1}{n}(\mathbf{E}(X^2) + (n-1)[\mathbf{E}(X)]^2) = \frac{1}{n}\sigma^2 + \mu^2,$$

implying

$$\mathbf{E}(\hat{\sigma}_2^2) = \sigma^2 + \mu^2 - (\frac{1}{n}\sigma^2 + \mu^2) = \frac{n-1}{n}\sigma^2.$$

So, to have unbiased estimate one has to take, instead of $\hat{\sigma}_2^2$, the estimate

$$\hat{\sigma}_3^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)^2.$$

Of course, for large n the difference between $\hat{\sigma}_2^2$ and $\hat{\sigma}_3^2$ disappears, and both these estimates are asymptotically effective.

Similarly, suppose we observe two sequences of independent identically distributed (i.i.d.) random variables $X_1, X_2, \ldots, Y_1, Y_2, \ldots$ distributed like X and Y respectively (let us stress that actually we observe the realizations of i.i.d. random vectors $(X_1, Y_1), (X_2, Y_2), \ldots$). An unbiased and asymptotically effective estimate for the covariance can be constructed as

$$\hat{Cov}(X, Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)(Y_i - \hat{\nu}_n),$$

where $\hat{\nu}_n = (Y_1 + \cdots + Y_n)/n$.

6.2.3 Waiting time paradox

As an illustration of an application of variance, let us discuss the so-called *waiting time paradox*.

One could be very annoyed by regularly waiting an hour for a bus at a bus stop, where buses are scheduled to run at 20-minute intervals. However, only if the buses run precisely at 20-minute intervals, your average waiting time (when you arrive at the stop at a random moment) will be 10 minutes. Of course this effect is relevant not only for buses, but for a variety of situations, when waiting times and queues are to be handled (i.e. supermarket cashpoints, internet sites, mobile phone networks, etc). The full development belongs to a highly applied domain of probability, called queueing theory. Here we shall only sketch a (not quite rigorous) argument leading to the calculation of the average waiting times.

To simplify the matter, let us assume that possible intervals between busses take only finite number of values, say $T_1, \dots, T_k$, with certain probabilities $p_1, \dots, p_k$. The average interval (usually posted on a timetable) is therefore

$$\mathbf{E}(T) = \sum_{j=1}^{k} p_j T_j.$$

We are interested in the average waiting time. Suppose first that all possible intervals follow each other periodically in a strictly prescribed order, say $T_1, T_2, \dots, T_k, T_1, T_2, \dots$ What will be your average waiting time, if you arrive at the stop at a random time t uniformly distributed on the interval $[0, T_1 + \cdots + T_k]$? If $T_1 + \cdots + T_{j-1} < t \leq T_1 + \cdots + T_j$ (i.e. your arrive at the jth period between buses), then you will wait for a time $W = T_1 + \cdots + T_j - t$. Taking expectation with respect to the uniform distribution of t yields

$$\begin{aligned}
&\mathbf{E}_{T_1,\dots,T_k}(W) \\
&= \frac{1}{T_1 + \cdots + T_k} \left(\int_0^{T_1} (T_1 - t)\, dt + \int_{T_1}^{T_1+T_2} (T_1 + T_2 - t)\, dt + \cdots \right. \\
&\quad \left. + \int_{T_1+\cdots+T_{k-1}}^{T_1+\cdots+T_k} (T_1 + \cdots + T_k - t)\, dt \right) \\
&= \frac{1}{T_1 + \cdots + T_k} \left(\int_0^{T_1} t\, dt + \int_0^{T_2} t\, dt + \cdots + \int_0^{T_k} t\, dt \right) \\
&= \frac{T_1^2 + \cdots + T_k^2}{2(T_1 + \cdots + T_k)}.
\end{aligned}$$

Similarly, if during a period of time there were m_1 intervals of length T_1, m_2 intervals of length $T_2, \dots, m_k$ intervals of length T_k (in any order), and you arrive uniformly randomly on this period, then your average waiting time will be

$$\mathbf{E}_{T_1,\dots,T_k;m_1,\dots,m_k}(W) = \frac{m_1 T_1^2 + \cdots + m_k T_k^2}{2(m_1 T_1 + \cdots + m_k T_k)} = \frac{q_1 T_1^2 + \cdots + q_k T_k^2}{2(q_1 T_1 + \cdots + q_k T_k)},$$

where $q_j = k_j/(k_1 + \cdots + k_m)$ denotes the frequencies of the appearance of the intervals T_j. But frequencies are equal approximately to probabilities (and approach them as the number of trials go to infinity), so that, approximately, if the intervals T_j occur with probabilities p_j, the above formula becomes

$$\mathbf{E}(W) = \frac{q_1 T_1^2 + \cdots + q_k T_k^2}{2(q_1 T_1 + \cdots + q_k T_k)} = \frac{\mathbf{E}(T^2)}{2\mathbf{E}(T)}, \tag{6.14}$$

which can be equivalently rewritten as

$$\mathbf{E}(W) = \frac{1}{2}\left[\mathbf{E}(T) + \frac{Var(T)}{\mathbf{E}(T)}\right]. \tag{6.15}$$

Hence, as we have expected, for a given average interval time $\mathbf{E}(T)$, the average waiting time can be be arbitrarily large depending on the variance of the interval lengths.

Similar arguments can be used in assessing traffic flows. For example, suppose n cars are driven along the race track, formed as a circumference of radius 1 km, with speeds $v_1, \ldots, v_n$ kilometers per hour, and a speed camera is placed at some point that registers the speeds of passing cars and then calculates the average (arithmetic mean) speed from all the observed ones. Would the average be $(v_1 + \cdots + v_n)/n$? Of course not! In fact, during a time T, the cars will cover $v_1 T, \ldots, v_n T$ circumferences respectively, so that the camera will register $(v_1 + \cdots + v_n)T$ cars with the average speed being

$$\frac{v_1^2 + \cdots + v_n^2}{v_1 + \cdots + v_n}.$$

If the speed of a car is a random variable with a given distribution, then this would turn to the expression $\mathbf{E}(V^2)/\mathbf{E}(V)$, which is similar to (6.14).

6.2.4 Hedging via futures

Futures and forwards represent contracts to buy or sell a commodity or an asset by a fixed price on a prescribed date in the future (the distinction between future and forwards will not be of any relevance to our elementary analysis). Futures markets can be used to hedge the risk, i.e. to neutralize it as far as possible.

Our plan here will be as follows: (1) we shall show how this kind of hedging works on a simple numerical example; (2) deduce the main formula for the optimal hedge ratio; (3) introduce the main idea (arbitrage) underlying the pricing of the futures. Thus we discuss on a simplest possible example the use of variance as a measure of risk and the basic idea of arbitrage for pricing financial securities. Both these ideas are central for financial economics, and will pop out further in many places.

Assume that on the 1st of April an oil producer negotiated a contract to sell 1 million barrels of crude oil on the 1st of August by the market price that will form on this latter date. The point is that this market price is unknown on the 1st of April. Suppose that on the 1st of April the crude oil futures price (on a certain Exchange) for August delivery is \$19 per barrel and each futures contract is for the delivery of 1000 barrels. The oil producer can hedge its risk by *shorting* 1000 futures contracts, i.e. by agreeing to sell $1000 \times 1000 = 10^6$ barrels on the 1st of August by the price of \$19 per barrel. Let us see what can then happen. Suppose the price for crude oil will go down and on the 1st of August become, say, \$18 per barrel. Then the company will realize \$18 million from its initial contract. On the other hand, the right to sell something by \$19 per barrel when the actual price is \$18 per barrel means effectively a gain of \$1 per barrel, i.e. the company will realize from its futures contract the sum of \$1 million. The total gain then will equal $18 + 1 = \$19$ million. Suppose now the opposite scenario takes place, namely, the price for crude oil goes up and on the 1st of August becomes, say, \$21 per barrel. Then the company will realize \$21 million from its initial contract. On the other hand, the obligation to sell something by \$19 per barrel when the actual price is \$21 per barrel means effectively the loss of \$2 per barrel, i.e. the company will loose from its futures contract the sum of \$2 million. The total gain then will equal $21 - 2 = \$19$ million. In both cases the total gain of the company is the one obtained by selling its oil according to the futures price for the August delivery. Thus the risk is totally eliminated.

Such a perfect hedging is possible because the commodity underlying a futures contract is the same as the commodity whose price is being hedged. However, this is not always possible, so that one can only hedge the risk using futures on a related commodity or an asset (say to hedge a contract on a certain product of crude oil by the futures on the oil itself). To assess (and to minimize) the risk correlations between these commodity prices should be then taken into account.

Suppose N_A units of assets are to be sold at time t_2 and a company is hedging its risk at time $t_1 < t_2$ by shorting futures contract on N_F units of a similar (but not identical asset). The *hedge ratio* is defined as

$$h = N_F/N_A.$$

Let S_1 and F_1 be the (known at time t_1) asset and futures prices at the initial time t_1 (let us stress that F_1 is the price, at which one can agree at time t_1 to deliver a unit of the related asset on the time t_2), and let S_2 and F_2 be the (unknown at time t_1) asset and futures prices at the time t_2 (F_2 is the price, at which one can agree at time t_2 to deliver a unit of the underlying asset at this time t_2, so it is basically equal to the asset price at time t_2). The gain of the company at time t_2 becomes

$$Y = S_2 N_A - (F_2 - F_1) N_F = [S_2 - h(F_2 - F_1)] N_A = [S_1 + (\Delta S - h\Delta F)] N_A,$$

where

$$\Delta S = S_2 - S_1, \quad \Delta F = F_2 - F_1.$$

We are aiming to minimize the risk, i.e. the spread of this random variable around its average. In other words we are aiming to minimize the variance $Var(Y)$. As S_1 and N_A are given, this is equivalent to minimizing

$$Var(\Delta S - h\Delta F).$$

Let σ_S, σ_F and ρ denote respectively the variance of ΔS, the variance of ΔF and the correlation between these random variables. Then

$$Var(\Delta S - h\Delta F) = (\sigma_S^2 + h^2\sigma_F^2 - 2h\rho\sigma_S\sigma_F).$$

Minimizing this quadratic expression with respect to h one obtains (check!) that the minimum is attained on the value

$$h^\star = \rho\frac{\sigma_S}{\sigma_F}, \tag{6.16}$$

which is the basic formula for the it optimal hedge ratio.

In practice, the coefficients σ_S, σ_F and ρ are calculated from the historical behavior of the prices using the standard statistical estimators (see Section 6.2.2).

Finally, let us give a rough idea on how futures contracts are priced. Let S_0 denote the price of the underlying asset at time t_0 and let r denotes the risk free rates (continuously compound), with which one can borrow and/or lend money (in practice rates r can be variable, but we only consider a simple situation with fixed rates). The futures price F to deliver this asset after a time T (called *time to maturity*) from the initial time t_0 should be equal to

$$F = S_0 e^{rT}. \tag{6.17}$$

Otherwise, *arbitrage opportunities* arise leading to a possibility of a free income and thus driving the price back to this level. In fact, suppose $F > S_0 e^{rT}$ (opposite situation is considered similarly). Then one can buy the asset by the price S_0 and short a future contract on it (i.e. enter into the agreement to sell the asset at time $t_0 + T$). Realizing this contract at time $t_0 + T$ would yield the sum F, which, when discounted to the present time, equals Fe^{-rT}. Since this is greater than S_0, one gets a free income.

6.2.5 Other measures of risk

Variance is definitely the simplest meaningful measure of risk. It is however non-adequate from several points of view. For instance, by increasing your possible winning, your risk can not be increased, but the variance can well do so. Looking for risks measures and analyzing their applicability remains an active area of research both in theory and in practice of finance.

We shortly mention here some basic alternatives. Apart from variance, one uses *semi-variance*

$$\int_{-\infty}^{\mu} (\mu - x)^2 f(x)dx = \mathbf{E}[(X - \mu)^2 \mathbf{1}_{X \le \mu}],$$

where $\mu = \mathbf{E}X$, *shortfall probability* $\mathbf{P}(X \le L)$ (that depends on an arbitrary parameter L), and possibly the most popular now is the Value at Risk, $VaR = VaR(X) = VaR_p(X)$, defined via the equation

$$\mathbf{P}(X \le -VaR) = p.$$

Thus losses exceeding VaR (the event $(X \le -VaR)$) can happen only with probability p. Often in banking practice $p = 0.01$ or 1%. As the disadvantages of VaR one can mention that (i) the event $(X \le -VaR)$ (and hence VaR itself) does not capture how bad things can really be, that is how large are the losses $X \le -VaR$; (ii) it depends on an arbitrary parameter p, which is effectively taken from nowhere, and so it can not be considered as an objective measure of risk.

Let us introduce some more recent trends in the analysis of risk. Let us say that a lottery is a gamble if it has positive expectation, but negative outcomes have nonzero probability. By a strategy let us mean a rule that, capital given, accepts or rejects a gamble with a distribution given. Any mapping Q from a set of gambles to $\mathbf{R}_+$ specifies a strategy S_Q: g is rejected at wealth w if $w < Q(g)$ and is accepted otherwise (i.e. $Q(g)$ is the minimal wealth at which g is accepted). Let us call strategies of type S_Q simple. Let us say that a strategy guarantees no-bankruptcy if for any sequence $G = \{g_1, g_2, \dots\}$ of gambles (that the strategy would reject or accept according to its rule and current capital w_t) and any initial wealth w_0, one has $\mathbf{P}(\lim_{t\to\infty} w_t = 0) = 0$.

Theorem 6.2.1 *(Foster–Hart, see [7]). There exists a unique mapping R from gambles to $\mathbf{R}_+$ such that the simple strategy S_Q guarantees no-bankruptcy $\iff Q(g) \geq R(g) \forall g$. Moreover, $R(g)$ solves the equation*

$$\mathbf{E} \log \left(1 + \frac{g}{R(g)}\right) = 0.$$

6.3 Optimal betting (Kelly's system) and money management

Suppose you have an edge in betting on a series of similar trials, i.e. the average gain in each bet is positive (say, as you might expect by trading on Forex using some advanced strategy). Then you naturally expect to win in a long run. However, if you are going to put all your bankroll on each bet, you would definitely loose instead. An obvious idea is therefore to bet on a fraction of your current capital. What is the optimal fraction?

To answer this question assume that you are betting on a series of Bernoulli trials with the probability of success p, when you get m times the amount you bet. And otherwise, with probability $1 - p$, you loose your bet. Consequently, if you invest one dollar, the expectation of your gain is mp dollars. This, assuming that you have an edge, is equivalent to assuming $mp > 1$, as we shall do now.

Let V_0 be your initial capital, and your strategy is to invest in each bet a fixed fraction α, $0 < \alpha < 1$, of your current bankroll. Let X_k denote the random variable that equals m, if the kth bet is winning, and equals zero otherwise, so that all X_k are independent and identically distributed. Hence after the first bet your capital becomes $V_1 = (1 - \alpha + \alpha X_1)V_0$, after the second bet it will be

$$V_2 = (1 - \alpha + \alpha X_2)(1 - \alpha + \alpha X_1)V_0,$$

and for any n your capital at time n will become

$$V_n = (1 - \alpha + \alpha X_n) \cdots (1 - \alpha + \alpha X_1)V_0.$$

We are aiming to maximize the ratio V_n/V_0, or equivalently its logarithm

$$\ln \frac{V_n}{V_0} = \ln(1 - \alpha + \alpha X_n) + \cdots + \ln(1 - \alpha + \alpha X_1).$$

Here the law of large numbers comes into play. Namely, according to this law, the average of the winning rates per bet

$$G_n = \frac{1}{n} \ln \frac{V_n}{V_0} = \frac{1}{n}[\ln(1 - \alpha + \alpha X_n) + \cdots + \ln(1 - \alpha + \alpha X_1)]$$

converges to the expectation

$$\begin{aligned}\phi(\alpha) &= \mathbf{E} \ln(1 - \alpha + \alpha X_1) \\ &= p \ln(1 - \alpha + \alpha m) + (1 - p) \ln(1 - \alpha).\end{aligned}$$

Therefore, to maximize the gain in a long run, one has to look for α that maximizes the function $\phi(\alpha)$. But this is an easy exercise in calculus yielding (check it!) that the maximum is attained at

$$\alpha^* = \frac{pm - 1}{m - 1}. \tag{6.18}$$

This is the final formula we aimed at, known as *Kelly's betting system*. This is not the full story, however, as we did not take into account the risk (we only used averages). A careful analysis (that we are not going into) allows one to conclude that by choosing the betting fraction α slightly smaller than α^* one can essentially reduce the risk with only a slight decrease of the expected gain. We refer to book [12] for a recent review of the theory and practice of Kelly's betting system.

Kelly's system reveals a distinguished role played by the logarithmic utility function in the theory of betting.

As an application, let us consider now an aspect of financial trading, namely *money management*, which by many active and successful traders is considered to be the most crucial one.

As a first (possibly rather artificial) example (taken from [16]), consider trading on a stock that each week goes up 80% or down 60% with equal probabilities 1/2. You clearly have an edge in this game, as the expectation of your gain is

$$\frac{1}{2}80\% - \frac{1}{2}60\% = 10\%.$$

On the other hand, suppose you invest (and reinvest) your capital for many weeks, say n times. Each time your capital is multiplied either by 1.8 (success) or by 0.4 (failure) with equal probabilities. During a large period of time, you can expect the number of successes and failures to be approximately equal (the law of large numbers!). Thus your initial capital V_0 would become

$$1.8^{n/2} 0.4^{n/2} V_0 = (0.72)^{n/2} V_0,$$

which quickly tends to zero as $n \to \infty$. So, do you really have an edge in this game?

The explanation to this paradoxical situation should be seen from the above discussion of Kelly's system. The point is that you should not invest all your capital at once (money management!).

To properly sort out the situation it is convenient to work in a more general setting. Namely, assume that the distribution of the return is given by a positive random variable R (i.e. in one period of investing your capital is multiplied by R) with a given distribution. Following the line of reasoning of the previous section, we can conclude that if you always invest the fraction α of your capital, then the average rate of return per investment would converge to

$$g(\alpha) = \mathbf{E} \ln(1 - \alpha + \alpha R).$$

We are interested in the maximum of this expression over all $\alpha \in [0, 1]$.

Proposition 6.3.1 Assume $\mathbf{E}(R) > 1$ (i.e. you have an edge in the game). Then (i) if $\mathbf{E}(1/R) \leq 1$ (your edge is overwhelming), then $\alpha = 1$ is optimal and your maximum average return per investment equals $g(1) = \mathbf{E} \ln(R)$; and (ii) if $\mathbf{E}(1/R) > 1$, then there exists a unique $\hat{\alpha} \in (0, 1)$ maximizing $g(\alpha)$, which is found from the equation

$$g'(\hat{\alpha}) = \mathbf{E} \frac{R - 1}{1 + (R - 1)\hat{\alpha}} = 0. \tag{6.19}$$

Proof Since the second derivative

$$g''(\alpha) = -\mathbf{E} \left(\frac{R - 1}{1 + (R - 1)\hat{\alpha}} \right)^2$$

is negative, the function g is concave (i.e. its derivative is decreasing), and there can be at most one solution of equation (6.19) that necessarily would specify a maximum value of g. The assumption $\mathbf{E}(R) > 1$ is equivalent to saying that $g'(0) > 0$. In case (i) one has $g'(1) \geq 0$, so that the solution $\hat{\alpha}$ can lie only to the right of $\alpha = 1$ and hence the maximum of $g(\alpha)$ is attained at $\alpha = 1$. In case (ii), $g'(1) < 0$, implying the existence of the unique solution to (6.19) in $(0, 1)$. □

Proposition 6.3.2 If R takes only two values m_1 and m_2 with equal probabilities (in the example at the beginning, $m_1 = 1.8$, $m_2 = 0.4$), the condition $\mathbf{E}(1/R) > 1$ rewrites as $m_1 + m_2 > 2m_1 m_2$, and if it holds, the optimal $\hat{\alpha}$ equals

$$\hat{\alpha} = \frac{1}{2(1-m_1)} + \frac{1}{2(1-m_2)}.$$

Proof Is left as an exercise. □

6.4 Portfolio, CAPM and factor models

6.4.1 Portfolio optimization

We introduce the Nobel price winning theory of the Markowitz mean-variance portfolio optimization.

Suppose $S_0^1, \dots, S_0^n$ are the initial prices of n securities, and let $S_T^1, \dots, S_T^n$ denote their prices at time T (unknown at the initial time). Let the returns be denoted by

$$1 + r_i = 1 + r_i(T) = S_T^i / S_0^i, \quad r = (r_1, \dots, r_n),$$

where r_i are *rates of return* (or *returns* for short). The main input data for the analysis, assumed to be known, are the expectations $\mathbf{E}(r_i) = \bar{r}_i$ and the covariance matrix Σ with entries $Cov(r_i, r_j) = \sigma_{ij}$. The initial capital is $y = y_0$.

The control parameters available to an investor are the numbers $(\phi_1, \dots, \phi_n)$ (not necessarily integers: liquidity!) of securities of each type satisfying the budget constraint:

$$\sum_{i=1}^n \phi_i S_0^i = y.$$

In terms of the *portfolio vector* $x = (x_1, \dots, x_n)$, $x_i = \phi_i S_0^i / y$, this constraint rewrites as $\sum_{i=1}^n x_i = 1$.

For a chosen portfolio vector, the wealth at time T becomes $y(T) = \sum_{i=1}^n \phi_i S_T^i$ with return

$$1 + r(x) = \frac{y(T)}{y} = \frac{1}{y} \sum_{i=1}^n \phi_i S_0^i (1 + r_i) = 1 + \sum_{i=1}^n x_i r_i,$$

$$\mathbf{E}(r(x)) = \sum_{i=1}^n x_i \bar{r}_i = (x, \bar{r}),$$

$$Var(r(x)) = \sigma^2(r(x)) = \sum_{i,j=1}^n x_i x_j \sigma_{ij} = (x, \Sigma x).$$

The two main problems suggested by Markowitz to describe an investor aiming at increasing the expected return and at the same time decreasing the risk measured in terms of the variance, can be formulated as follows: (i) find $\min_x \sigma^2(r(x))$ with $\mathbf{E}(r(x))$ given or bounded below by a given constant R; (ii) find $\max_x \mathbf{E}(r(x))$ with $\sigma^2(r(x))$ given.

As seen from the analysis below, these problems are in some sense equivalent. Let us analyze the first one using the method of Lagrange multipliers. The Lagrange function can be taken as (the coefficient $\frac{1}{2}$ is inserted for convenience):

$$L = \frac{1}{2}(x, \Sigma x) - \lambda((x, \bar{r}) - R) - \mu(\sum_{i=1}^{n} x_i - 1).$$

The equations for critical points

$$\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial \lambda} = 0, \quad \frac{\partial L}{\partial \mu} = 0$$

yield the system ($\mathbf{1}$ is the vector with all coordinates one) of equations

$$\Sigma x = \lambda \bar{r} + \mu \mathbf{1}, \quad (x, \bar{r}) = R, \quad (x, \mathbf{1}) = 1. \tag{6.20}$$

From the first equation

$$x = \Sigma^{-1}(\lambda \bar{r} + \mu \mathbf{1}). \tag{6.21}$$

The last two equations in (6.21) yield the linear system

$$\begin{cases} A\mu + B\lambda = 1, \\ B\mu + C\lambda = R, \end{cases} \tag{6.22}$$

with constant coefficients

$$A = (\mathbf{1}, \Sigma^{-1}\mathbf{1}), \quad B = (\bar{r}, \Sigma^{-1}\mathbf{1}), \quad C = (\bar{r}, \Sigma^{-1}\bar{r}). \tag{6.23}$$

Hence

$$\lambda = \lambda(R) = \frac{AR - B}{AC - B^2}, \quad \mu = \mu(R) = \frac{C - BR}{AC - B^2}. \tag{6.24}$$

By the Cauchy inequality (applied to quadratic form $(x, \Sigma^{-1}x)$), $AC \geq B^2$, and $AC = B^2 \iff \bar{r}$ is proportional to $\mathbf{1}$. Denote $x^* = x^*(R)$ the corresponding vector (6.21). Then $\Sigma x^* = \lambda \bar{r} + \mu$ with λ, μ from (6.24). For $\sigma = \sigma(r(x^*(R)))$ we get

$$\sigma^2 = Var(r(x^*)) = (x^*, \Sigma x^*) = \lambda(x^*, \bar{r}) + \mu(x^*, \mathbf{1})$$

$$= \lambda R + \mu = \frac{AR^2 - 2BR + C}{AC - B^2} = \frac{A(R - B/A)^2}{AC - B^2} + \frac{1}{A}.$$

Hence $\sigma_m^2 = \min_R Var(r(x^*)) = 1/A$, the minimum is attained at $R_m = B/A$ with $\lambda_{\min} = 0$, and the last equation rewrites as

$$\frac{\sigma^2}{\sigma_m^2} - \frac{A}{AC - B^2} \frac{(R - R_m)^2}{\sigma_m^2} = 1. \tag{6.25}$$

These optimal returns and their deviations R, σ lie on a hyperbola. The upper part of this hyperbola is called (for obvious reasons) the *efficient frontier* (see Fig. 6.4).

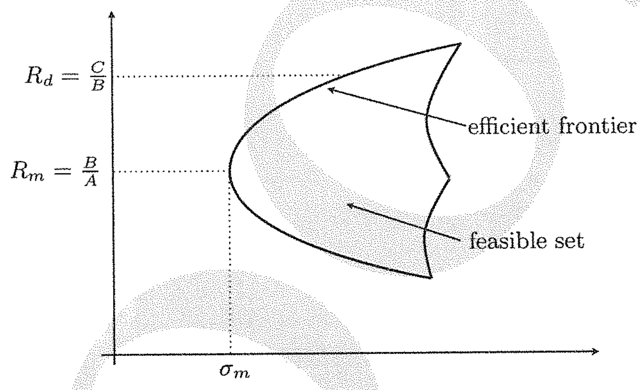

Figure 6.4 Efficient frontier and two funds.

Two feasible sets can be naturally considered corresponding to two reasonable posits of the problem: one with shorting not permitted (that is, all x_i non-negative), and another with shorting allowed (without the above constraint).

Two special points on the hyperbola (6.25) are worth singling out, one with vanishing λ and one with vanishing μ, the corresponding x^* will be denoted x_g and x_d:

(1) $\lambda = 0 \Longleftrightarrow R = R_m = \dfrac{B}{A} \Longrightarrow \mu = \dfrac{1}{A} = \sigma_m^2, \quad x^* = x_g = \dfrac{1}{A}\Sigma^{-1}\mathbf{1},$

(2) $\mu = 0 \Longleftrightarrow R = \dfrac{C}{B} \Longrightarrow \lambda = \dfrac{1}{B}, \quad x^* = x_d = \dfrac{1}{B}\Sigma^{-1}\bar{r}.$

Notice $\frac{C}{B} \geq \frac{B}{A} = R_m$.

For any R we have

$$x^* = \lambda B x_d + \mu A x_g,$$

leading to the following crucial result:

Theorem 6.4.1 The two-fund theorem. *There exist two efficient portfolios, namely x_d and x_g, such that any efficient portfolio is their linear combination.*

6.4.2 Portfolio optimization with a risk-free asset

Additionally to the above setting, suppose there exists a risk-free asset with the rate of growth r_f and some initial price S_0^0. Let $x = (x_1, \dots, x_n)$ still denote the part held on risky assets. Then $\sum_{i=1}^n x_i \le 1$ and $x_0 = 1 - (x, \mathbf{1}) = \phi_0 S_0^0 / y_0$ is the part held on the risk-free asset.

The main problem can now be formulated as the problem of finding the minimum of $Var(r(x))/2$ subject to $(x, \bar{r}) + (1 - (x, \mathbf{1}))r_f = R$. The Lagrange function is

$$L = \frac{1}{2}(x, \Sigma x) - \lambda[(x, \bar{r}) + (1 - (x, \mathbf{1}))r_f - R].$$

The optimality conditions are

$$\Sigma x = \lambda(\bar{r} - r_f \mathbf{1}), \quad (x, \bar{r} - r_f \mathbf{1}) = R - r_f. \tag{6.26}$$

Thus, for a solution pair (x^*, λ)

$$x^* = \lambda \Sigma^{-1}(\bar{r} - r_f \mathbf{1}), \tag{6.27}$$

$$\lambda = \frac{R - r_f}{(\bar{r} - r_f \mathbf{1}, \Sigma^{-1}(\bar{r} - r_f \mathbf{1}))} = \frac{R - r_f}{C - 2r_f B + r_f^2 A}. \tag{6.28}$$

Then

$$x^* = \frac{R - r_f}{C - 2r_f B + r_f^2 A} \Sigma^{-1}(\bar{r} - r_f \mathbf{1}), \tag{6.29}$$

and

$$\begin{aligned} \sigma^2 &= \sigma^2(r(x^*)) = (x^*, \Sigma x^*) \\ &= \lambda^2(\bar{r} - r_f \mathbf{1}, \Sigma^{-1}(\bar{r} - r_f \mathbf{1})) = \frac{(R - r_f)^2}{C - 2r_f B + r_f^2 A}, \end{aligned} \tag{6.30}$$

or

$$R = r_f \pm \sigma \sqrt{C - 2r_f B + r_f^2 A}, \tag{6.31}$$

which effectively decomposes into two linear functions. The upper part is called the *efficient frontier* or *efficient line* and has the slope

$$\tan \theta = \sqrt{C - 2r_f B + r_f^2 A}. \tag{6.32}$$

The overall result of the analysis is quite impressive: the totality of data (of order n^2 for n assets) is reduced to one line!

The *tangent portfolio* is defined as the portfolio on the efficient line such that $(x_T, \mathbf{1}) = 1$, that is, the unique efficient portfolio without risk-free investment. Denote $r_T = (x_T, r)$, $\bar{r}_T = (x_T, \bar{r})$ its return and σ_T its deviation. The efficient line (see Fig. 6.5) on (R, σ) plane is then

$$R = r_f + \frac{\bar{r}_T - r_f}{\sigma_T}\sigma.$$

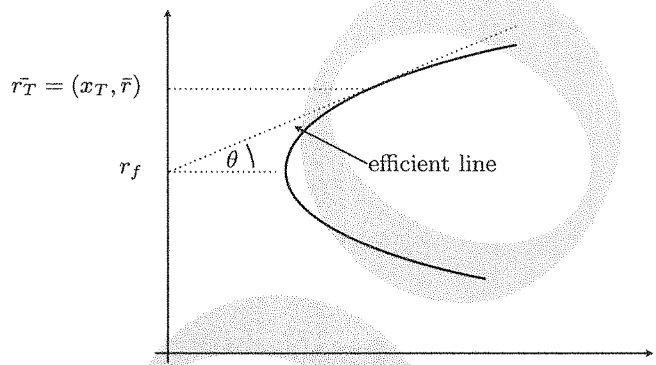

Figure 6.5 Efficient line and tangent portfolio.

For the tangent portfolio we have

$$(x_T, \mathbf{1}) = 1 = \lambda_T[(\mathbf{1}, \Sigma^{-1}\bar{r}) - r_f(1, \Sigma^{-1}\mathbf{1})],$$

implying

$$\lambda_T = \frac{1}{B - r_f A}, \quad x_T = \frac{1}{B - r_f A}\Sigma^{-1}(\bar{r} - r_f\mathbf{1}), \quad (x_T, \bar{r}) = \frac{C - r_f B}{B - r_f A}. \tag{6.33}$$

Hence for any R

$$x^* = \lambda\Sigma^{-1}(\bar{r} - r_f\mathbf{1}) = \lambda(B - r_f A)x_T, \tag{6.34}$$

leading to the following crucial result:

Theorem 6.4.2 The one-fund theorem. *There exists a portfolio of risky assets, namely x_T: any efficient portfolio can be constructed as a combination of the fund x_T and some risk-free assets.*

Let us find the standard deviation of the tangent portfolio x_T. From (6.30) and the first equation of (6.33),

$$\sigma_T^2 = \lambda_T^2(\bar{r} - r_f\mathbf{1}, \Sigma^{-1}(\bar{r} - r_f\mathbf{1})) = \frac{C - 2r_f B + r_f^2 A}{(B - r_f A)^2}, \tag{6.35}$$

and using the last equation in (6.33),

$$\sigma_T^2 = \frac{1}{B - r_f A}\left[(x_T, \bar{r}) + \frac{r_f(r_f A - B)}{B - r_f A}\right] = \frac{\bar{r}_T - r_f}{B - r_f A} = \lambda_T(\bar{r}_T - r_f). \tag{6.36}$$

Let us now use (6.27) in the opposite direction:

$$\bar{r} - r_f\mathbf{1} = \frac{1}{\lambda_T}\Sigma x_T. \tag{6.37}$$

By the definition of Σ, $(\Sigma x)_i = Cov(r_i, (r, x))$ for any x. Hence, by (6.36) and (6.37),

$$\bar{r} - r_f\mathbf{1} = \frac{Cov(r, r_T)}{\sigma_T^2}(\bar{r}_T - r_f). \tag{6.38}$$

Thus we have proved the following:

Theorem 6.4.3 *The average returns of the initial assets are expressed in terms of the tangent portfolio by formula (6.38).*

We say that an investor with portfolio x^* *lends money to the market* if the amount put on the risk-free asset is positive, that is $(x^*, \mathbf{1}) < 1$, or equivalently (using (6.34))

$$(x^*, \mathbf{1}) = \lambda(B - r_f A)(x_T, \mathbf{1}) = \lambda(B - r_f A) < 1,$$

which in turn is equivalent to $\lambda < \lambda_T$ or

$$(x^*, \bar{r}) = \lambda(B - r_f A)(x_T, \bar{r}) < (x_T, \bar{r}) = \bar{r}_T.$$

Otherwise he/she borrows money.

6.4.3 Capital Asset Pricing Model (CAPM)

The main conjecture of the *Capital Asset Pricing Model* (CAPM) can be formulated by stating that the totality of all assets on the market is an efficient portfolio, and hence the tangent portfolio (as it does not include risk-free assets, by definition). This universal tangent portfolio is then referred to as the *market portfolio*. An asset's *weight* in this portfolio is then naturally defined as the proportion of the asset's capital value to the total market capital value. Briefly, the CAPM states that *the market portfolio is efficient.*

Thus the tangent portfolio x_T is now the collection of all assets of the market x_M. The efficient line is then called the *capital market line*:

$$R = r_f + \frac{\bar{r}_M - r_f}{\sigma_M}\sigma. \tag{6.39}$$

Its slope

$$\lambda_M = \frac{\bar{r}_M - r_f}{\sigma_M}$$

is called the *market price of risk.*

The fundamental equation (6.39) can be formulated as the following 'equation in words':

Expected return = Price of time + (Price of risk) × (Amount of risk).

Practically, for choosing an appropriate investment on the efficient line, one uses an individual utility function that somehow measures an investor's aversion to risk. The simplest class of utility functions is of course quadratic, leading to the following optimization problem: for the utility function $R - c\sigma^2$ find the optimal point (maximizing the utility) on the capital line. Here c is a constant specifying the risk aversion of an investor. The inverse number, $1/c$, is referred to as *risk tolerance.*

The main analytic result of CAPM represents a direct consequence of Theorem 6.4.3, which can be formulated as follows.

Theorem 6.4.4 CAPM: main theorem. *If the market portfolio M is efficient (i.e. the CAPM conjecture holds), the expected return of any asset satisfies*

$$\bar{r}_i - r_f = \beta_i(\bar{r}_M - r_f), \quad \beta_i = \frac{\sigma_{iM}}{\sigma_M^2}, \quad \sigma_{iM} = Cov(r_i, r_M). \tag{6.40}$$

Function (6.40), as a function of β_i, or the corresponding line in the (R, β)-plane, is called the *security market line* (SML). Alternatively, one represents SML as a line in the (R, σ_{iM})-plane:

$$\bar{r}_i = r_f + \frac{\lambda_M}{\sigma_M}\sigma_{iM}.$$

Remark 6.4.5 Some authors refer to $\lambda'_M = (\bar{r}_M - r_f)/\sigma_M^2$ as the market price of risk, so that CAPM in the (R, σ_{iM})-plane becomes $\bar{r}_i = r_f + \lambda'_M\sigma_{iM}$.

As a simple numeric example, let us assume that $r_f = 8\%, \bar{r}_M =$