Chapter Seven

Decision-Making under Uncertainty: The Static Case

Uncertainty is the rule in most financial decision-making problems. The prototypical case is the allocation of wealth to a set of assets with uncertain returns. If we make a here-and-now decision and observe the return of the portfolio after a given holding period, we are considering a **static** decision problem, since we disregard the possibility of adjusting our decisions along the way, when we observe the actual unfolding of uncertain risk factors. This is not to say that, in reality, the portfolio will not be adjusted after a while, possibly by solving the same model again; the point is that this is not explicitly considered in the decision model itself. On the contrary, **multistage** decision models take into account the possibility of updating decisions, depending on the incoming information flow over time. It is important to avoid a potential confusion between *multistage* and *multiperiod* models. A multiperiod problem requires the planning of decisions to be executed over a sequence of time instants. However, if the plan is specified here and now, once for all, the problem is actually static, as there is no dynamic adaptation. The solution of a multiperiod problem is a sequence of numbers, representing the decisions that are supposed to be implemented, no matter what. On the contrary, the solution of a multistage problem consists of a set of random variables, since decisions will be contingent on the realization of uncertain states. We may also explicitly express decisions as functions of the uncertain states or, alternatively, as functions of the realization of random risk factors.

In this chapter, we lay down the conceptual foundations of decision-making under uncertainty in the static, single-period case. It is useful to consider portfolio decisions as an application framework to understand the related issues; however, what we describe here is also relevant for asset pricing. Here, we do not consider either model building or solution algorithms.[1] In Chapter 8, we discuss a specific relevant case in detail, mean–variance portfolio optimization, whereas in Chapter 15 we outline more advanced models, including multistage

[1]A very simple introduction to deterministic optimization models and solution algorithms is given in Chapter 12 of [4]. Chapter 13 therein describes some models for decision-making under uncertainty, including, but not limited to financial problems.

ones. Here, we introduce three possible approaches to decision-making under uncertainty that are relevant to finance:

- Utility functions
- Mean–risk models
- Stochastic dominance

We start with a few simple introductory examples in Section 7.1. Then, in Section 7.2, we show that financial decision-making cannot rely on simple maximization of expected wealth or expected return. Risk should be carefully accounted for. One way for doing so, albeit not quite a practical one, is by introducing expected utility, as we illustrate in Section 7.3. A more practical approach is based on the definition of suitable risk measures and the solution of a mean–risk optimization problem. Mean–risk models, as we show in Section 7.4, are the foundation of the ubiquitous mean–variance portfolio optimization framework. However, there is no reason why we could not replace standard deviation (or variance) of return by an alternative risk measure. We have already introduced value-at-risk in Section 2.2.2. Here, we discuss some basic properties that a *coherent* risk measure should satisfy. As it turns out, value-at-risk is not quite satisfactory in this respect. Then, in Section 7.5, we outline a third approach, stochastic dominance. This last section is included for the sake of completeness, but it is not needed for the remainder of this book and may be safely skipped. We also include a couple of theorem proofs in Supplement S7.1, which may be safely skipped, too. Usually, we do not include complete and overly rigorous proofs, given the introductory nature of this book. However, some of them may be instructive and useful to the interested reader.

7.1 Introductory examples

A couple of simple examples may help in framing the kind of problems that we want to tackle in this chapter.

Example 7.1 A choice among lotteries

Consider the choice among the four lotteries depicted in Table 7.1. These lotteries are characterized by uncertain payoffs, which we model by four discrete random variables $L_i(\omega)$, $i = 1, 2, 3, 4$, taking values corresponding to three equally likely outcomes ω_1, ω_2, and ω_3. For each $L_i(\omega)$, in the table we also report its expected value μ_i and standard deviation σ_i. Which lottery should we choose?

It is easy to see that lottery L_4 would not be chosen, since its payoff is dominated by L_1 (as well as by L_3):

$$L_4(\omega_k) \le L_1(\omega_k), \qquad k = 1, 2, 3,$$

Table 7.1 **Choice among four lotteries.**

Lottery	ω_1	ω_2	ω_3	μ_i	σ_i
$L_1(\omega)$	100	200	300	200	81.65
$L_2(\omega)$	-800	200	1200	200	816.50
$L_3(\omega)$	150	200	244	198	38.40
$L_4(\omega)$	100	200	150	150	40.82

with strict inequality in scenario ω_3. Lottery L_2 is obtained from L_1 by shifting a payoff of 900 units from scenario ω_1 to ω_3. Thus, we do not change the expected value, but the payoff has a much larger variability, as measured by the standard deviation. Many would agree that, since the expected value is the same and there is less uncertainty, lottery L_1 should be preferred to L_2.

Actually, this is a matter of individual taste and depends on how much we like or dislike taking risk. If we are risk-averse, chances are that we may even like L_3 the most. This lottery is obtained from L_1 by increasing the payoff for event ω_1 by 50 and decreasing the payoff for ω_3 by 56. Thus, the expected value μ_3 is only 198, but the standard deviation is considerably reduced.

In Example 7.1, we have only considered expected value and standard deviation of a lottery. Indeed, there is a large body of knowledge, broadly referred to as **modern portfolio theory**, which revolves around this view. However, this may not quite enough. As we said, if we compare the payoffs of lotteries L_1 and L_4 in Table 7.1, state by state, the latter is clearly dominated. However, we cannot reach a clear conclusion by just considering expected value and standard deviation of the two payoffs, since $\mu_4 < \mu_1$ and $\sigma_1 > \sigma_4$. One issue is that standard deviation does not capture the features of a very skewed random variable, associated with an asymmetric probability distribution. Example 7.2 below further illustrates this point.

Example 7.2 A dominated lottery

Let us consider the two lotteries described in Table 7.2. Note that the states of nature (outcomes) are not equiprobable. We find the expected value and the standard deviation of the payoff of lottery $L_1(\omega)$

Table 7.2 **A dominated lottery.**

State	ω_1	ω_2	ω_3
Probability	0.4	0.4	0.2
Payoff $L_1(\omega)$	10	50	100
Payoff $L_2(\omega)$	10	50	500

as follows:

$$\mu_1 = 0.4 \times 10 + 0.4 \times 50 + 0.2 \times 100 = 44$$
$$\sigma_1 = \sqrt{0.4 \times 10^2 + 0.4 \times 50^2 + 0.2 \times 100^2 - 44^2} \approx 33.23$$

By the same token, for lottery $L_2(\omega)$ we find $\mu_2 = 124$ and $\sigma_2 \approx 188.85$. If we compare the two alternatives in terms of expected value and standard deviation, there is an unclear tradeoff between the two lotteries, as the second one is more attractive in terms of expected payoff, but it looks riskier. However, if we compare the payoffs state by state, $L_1(\omega)$ is clearly dominated by $L_2(\omega)$. The problem is that the large payoff of lottery $L_2(\omega)$ in state ω_3 increases not only the expected value, but also standard deviation. Its distribution is positively skewed, and a symmetric deviation measure, like standard deviation, does not properly account for this feature. We should also notice that, if we introduce a negative skew, standard deviation will not tell the difference with respect to a corresponding positive skew.

We have considered simple lotteries that may be represented by a *discrete* random variable X that takes values x_j with probabilities p_j, corresponding to scenarios (also called outcomes or states of the world) ω_j, $j = 1, \ldots, m$. In risk management, this random variable usually represents loss, rather than profit, return, or payoff. We may also consider *continuous* random variables, as is common in asset allocation problems.

Example 7.3 Static asset allocation

We are endowed with wealth W_0 that we should allocate among a set of n assets with current price S_{i0}, $i = 1, \ldots, n$. At the end of a holding period of length T, the prices of the assets are represented by continuous random variables $S_{iT}(\omega)$. If we assume that assets are infinitely divisible and short-selling is not allowed, our decision can be represented by decision variables $h_i \geq 0$, $i = 1, \ldots, n$, corresponding to the holding of each asset, i.e., the number of stock shares of firm i included in the portfolio.

Decision variables are subject to a budget constraint,

$$\sum_{i=1}^{n} h_i S_{i0} = W_0,$$

and define a random variable,

$$W_T(\omega) = \sum_{i=1}^{n} h_i S_{iT}(\omega),$$

which is the random terminal wealth for each outcome $\omega \in \Omega$.

In this case, the problem does not just require ranking a few simple lotteries. By choosing the portfolio holdings we define a continuous probability distribution of terminal wealth, and we might choose the most preferred one by defining and optimizing a suitable **functional** $F(\cdot)$, mapping a random variable into the set of real numbers:

$$\max_{h_1,\ldots,h_n} F\big[W_T(\omega)\big].$$

We are talking about a functional rather than a function, since we are mapping random variables (which are function themselves, and not just numerical variables), to real numbers. If we can find a suitable functional $F(\cdot)$, we may map a possibly complicated preference structure into the simple ordering of real numbers.

Throughout the chapter, we assume that we have a credible stochastic characterization of the probability distribution of uncertain risk factors. The distribution may be considered as an **objective** assessment of uncertainty, but it is most likely to be at least partially **subjective**. The difference is that, ideally, all market participants should agree on a truly objective representation of uncertainty. On the contrary, market views are to some extent subjective. In more sophisticated models, we explicitly consider distributional ambiguity and look for a *robust* solution. In such a case, we could be uncertain about a set of plausible probability distributions, or we might even take a radical view and give up the idea of a stochastic representation of uncertainty. In this chapter, for the sake of simplicity, we assume that a reliable stochastic representation of uncertainty is available.

7.2 Should we just consider expected values of returns and monetary outcomes?

Whenever we bet money on a lottery or invest wealth in risky assets, we pay due attention to the expected value of the payoff, i.e., a monetary outcome, or to expected return. The expected value is quite likely to be the first feature we consider, when dealing with a probability distribution. However, let us ask the following questions:

- Given a set of assets or alternative financial portfolios, should we just select the one with the largest expected return? No doubt, this would make life much easier when dealing with decision-making under uncertainty. However, as we show below, this does not take risk into account and may lead to quite unreasonable decisions. As a general rule, larger expected returns come with a larger exposure to risk, and this leads to the need of assessing difficult risk–return tradeoffs.
- Should we consider an asset with *negative* expected return for inclusion within a portfolio? Even if we suspect that expected return does not tell the whole story, one is tempted to think that there is little good to be expected from such an asset, unless short-selling is allowed. However, this simplistic view does not consider the correlations between returns. An asset with a negative expected return may be negatively correlated with other assets and contribute to reducing risk. Derivatives such as futures and forward contracts are in fact included in an asset portfolio (possibly a nonfinancial one, involving commodities) to reduce risk by exploiting a negative correlation. In real life, indeed, we often purchase insurance, which is an asset with (hopefully) negative expected return, as we expect to pay the insurance premium but hope that a severe accident will not occur.[2]
- Given a financial asset with an array of random payoffs, can we just consider the expected value of the payoff to price the asset fairly? This is a relevant question, when dealing with derivatives and insurance contracts. If an insurance company faces a large set of small-scale and independent risks, it may be argued that finding the *actuarially fair* price of an insurance policy, by estimating the expected cash outflow for the company, may be a good strategy. However, this need not apply in general, and the insurance business can get quite dangerous when risks turn out to be correlated.[3]

Among other things, these questions show the link between the three basic problems of asset allocation, risk management, and asset pricing.

[2] As a further, but quite different example, lottery tickets have a negative expected payoff, since it is unlikely that we will win. So, it seems that we may be risk lovers, at least in the small.

[3] A good lesson in this respect comes from the default risks on mortgages in 2008, leading to the subprime crisis and the ultimate demise of Lehman Brothers.

7.2.1 FORMALIZING STATIC DECISION-MAKING UNDER UNCERTAINTY

In this section, we consider possible ways of formalizing a static problem under uncertainty. A generic optimization model may be written as

$$\min_{\mathbf{x}\in S} f(\mathbf{x}),$$

where:

- $\mathbf{x} \in \mathbb{R}^n$ is the vector of decision variables
- $S \subseteq \mathbb{R}^n$ is the set of feasible solutions
- $f(\cdot)$ is the objective function, mapping solutions (vectors in $\mathbb{R}^n$) into a numerical evaluation of their quality (a number in $\mathbb{R}$)

In finance, the objective function is likely to be related to a monetary outcome, like profit/loss, or to a return. Depending on the choice of the objective function, the problem may be a minimization or a maximization one. In most fields of practical interest, some data or parameters of the optimization model are uncertain. One way of stating this is by considering a vector of random variables $\boldsymbol{\xi}(\omega)$, where $\omega \in \Omega$ is a random outcome, corresponding to a scenario, within the sample space Ω. Then, the objective function becomes a function $f\big(\mathbf{x}, \boldsymbol{\xi}(\omega)\big)$ of both controllable and uncontrollable variables, and the feasible set may be random, too. This may have two consequences:

- The quality of the solution that we find is random and may turn out to be not quite what we expect.
- Possibly worse, the solution may even turn out to be infeasible for some realizations of the random data.

As we have pointed out before, in Example 7.3, a specific choice $\mathbf{x}_0$ of the decision variables defines the distribution of a random variable $Y_0 = f\big(\mathbf{x}_0, \boldsymbol{\xi}(\omega)\big)$, and we need a way to rank probability distributions. The simplest choice is to rank distributions by the corresponding expected value. Thus, we might consider an optimization problem like

$$\min_{\mathbf{x}\in S} \mathbb{E}\big[f\big(\mathbf{x}, \boldsymbol{\xi}(\omega)\big)\big].$$

However, just taking the expected value of an objective like cost or profit may not account for different attitudes toward risk. Thus, in general, we may consider a transformation of the random performance measure $f\big(\mathbf{x}, \boldsymbol{\xi}(\omega)\big)$, say, $\mathcal{R}_0\big[f\big(\mathbf{x}, \boldsymbol{\xi}(\omega)\big)\big]$, which should be considered as a **risk functional**. In concrete, as we shall see later, we may consider utility functions or mean–risk models.

Actually, stating an optimization problem under uncertainty in a precise way is not quite trivial, as different approaches may be pursued to model the interplay of decisions and observations, i.e., how to define a dynamic decision strategy, as well as how to cope with potential infeasibility of decisions made before knowing the values of uncertain data. To be more concrete, let us assume that the feasible set is explicitly described by a set of inequalities:

$$g_i\big(\mathbf{x}, \boldsymbol{\xi}(\omega)\big) \leq 0, \qquad i = 1, \ldots, m.$$

Clearly, for a given $\mathbf{x}$, we cannot be sure that the inequality will be satisfied for every value of $\boldsymbol{\xi}$. If we insist on guaranteed feasibility in every scenario,[4] an overly fat solution may be obtained. Here, too, we may introduce functionals $\mathcal{R}_i$, $i = 1, \ldots, m$, and require

$$\mathcal{R}_i\big[g_i(\mathbf{x}, \boldsymbol{\xi}(\omega))\big] \leq 0, \qquad i = 1, \ldots, m.$$

A naive approach would be to require that the expected value of the constraint function g_i is negative or zero, but this would be a very weak statement of an uncertain constraint. To see why, consider a standard normal distribution, where the expected value is zero, but the probability of a strictly positive value is 50%. As an alternative, we may settle for a probabilistic satisfaction of the constraints. We may introduce a set of **individual** chance constraints,

$$\mathrm{P}\big\{g_i\big(\mathbf{x}, \boldsymbol{\xi}(\omega)\big) \leq 0\big\} \geq 1 - \alpha_i, \qquad i = 1, \ldots, m,$$

or a **joint** chance constraint,

$$\mathrm{P}\big\{g_i\big(\mathbf{x}, \boldsymbol{\xi}(\omega)\big) \leq 0, \quad i = 1, \ldots, m\big\} \geq 1 - \alpha.$$

We should ask whether chance constraints are a suitable modeling framework, which means: (a) whether they allow us to express a financial decision-making problem in a sensible way, and (b) whether they lead to model formulations that may be efficiently solved. We will discuss this matter in Section 15.6.1.

We note again that, in a static decision problem under uncertainty, the solution is not dynamically adapted according to contingencies. Furthermore, it is practically impossible to find feasible solutions to problems involving random equality constraints. To this aim, we may take advantage of a more flexible modeling framework, stochastic programming with recourse, which will be introduced in Chapter 15.

7.2.2 THE FLAW OF AVERAGES

In common wisdom, we often consider loose statements of the law of large numbers, which is typically referred to as the "law of averages." Here, we rather consider the *flaw* of averages.[5] A comparison of the lotteries in Table 7.1 suggests that ranking alternatives on the basis of the expected value is probably neither safe nor sensible. Let us consider a few further examples reinforcing the point.

[4]Technically, we say that constraints are satisfied *almost surely*, i.e., with the exception of a set of null measure. Alternatively, we say that constraints are satisfied with probability one. When dealing with a finite set Ω of discrete outcomes, this boils down to the satisfaction of constraints in every discrete scenario.

[5]See [16].

Example 7.4 A single bet vs. multiple repeated bets

Consider a simple lottery based on the flip of a fair coin: If it lands tails, we win €10, otherwise we lose €5. Should we play this lottery? The expected payoff is €2.5, and most people answer that they would be willing to take the gamble. If we spice things up and scale the payoff by a factor of one *million*, the answer turns probably negative. Sure, an expected payoff of €2.5 million is quite palatable, but the considerable risk of losing €5 million makes the gamble not attractive to most people.

However, imagine playing the gamble repeatedly many times, say, one thousand times. Our answer could change if we are allowed to settle the score at the end of the game. Let X_i be the payoff of flip number i, $i = 1, \ldots, n$, where n is the number of independent and identically distributed flips. Thus, the variables X_i are i.i.d. random variables. Let $Y = \sum_{i=1}^{n} X_i$ be the total payoff, and let us denote the common expected value and standard deviation of the variables X_i by μ_X and σ_X, respectively. Then, the **coefficient of variation** of Y, under the hypothesis of independent flips, is

$$C_Y \doteq \frac{\sigma_Y}{|\mu_Y|} = \frac{\sqrt{n}\sigma_X}{n\mu_X} = \frac{C_X}{\sqrt{n}}, \tag{7.1}$$

where we assume $\mu_X > 0$. If n is large, the expected overall payoff becomes virtually certain (this is an informal glimpse of the law of large numbers). However, if we may go bankrupt along the way (i.e., we settle each flip of the coin individually, rather than assessing the overall profit/loss at the end of the game) or risks are correlated, this is not true anymore.

The natural interpretation of Example 7.4 is in terms of a bet repeated over time. An alternative view, which is relevant to insurers, concerns multiple bets taken at the same time. Indeed, Eq. (7.1) shows why an insurer providing coverage for a large number of uncorrelated risks may rely of expectations, plus some fudge consisting of reserves. However, correlated risks are much more dangerous. For an insurer, the random variables X_i correspond to losses. Let us assume that losses are pairwise correlated in the same way and that the common correlation coefficient is ρ. Then, the variance of the total loss becomes

$$\begin{aligned}
\mathrm{Var}(Y) &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathrm{Cov}(X_i, X_j) \\
&= n\sigma_X^2 + n \cdot (n-1)\rho\sigma_X^2 \\
&= n\sigma_X^2 \cdot \big[1 + (n-1)\rho\big].
\end{aligned}$$

In the limit case $\rho = 1$, we have $\mathrm{Var}(Y) = n^2\sigma_X^2 = \mathrm{Var}(nX)$ and there is no diversification of risk, in the sense that the coefficient of variation becomes

$$C_Y = \frac{n\sigma_X}{n\mu_X} = C_X.$$

As an example of correlated risks, we may think of home insurance in a region prone to earthquakes, or mortgage defaults under economic recession, as it happened during the subprime mortgage crisis.

Example 7.5 **Putting all of our eggs in one basket**

Consider an investor who must allocate her wealth to n assets. The return of each asset, indexed by $i = 1, \ldots, n$, is a random variable R_i with expected value $\mu_i = \mathbb{E}[R_i]$. Asset allocations may be expressed by decision variables w_i, representing the fraction of wealth invested in asset i. If we rule out short-selling, these decision variables are naturally bounded by $0 \le w_i \le 1$. If we assume that the investor should just maximize expected return, she should solve the problem

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} \mu_i w_i \\ \text{s.t.} \quad & \sum_{i=1}^{n} w_i = 1 \\ & w_i \ge 0. \end{aligned}$$

This is a simple model that we have already met in Section 2.1.1, Eq. (2.1), and we know its quite trivial solution: Just pick the asset with maximum expected return, $i^* = \arg\max_{i=1,\ldots,n} \mu_i$, and set $w_{i^*} = 1$. It is easy to see that this concentrated portfolio is a very dangerous bet. In practice, portfolios are diversified, which means that decisions depend on something beyond expected values. Furthermore, one would also include additional constraints on portfolio composition, bounding exposure to certain geographic areas or types of industry, and they would render the above trivial solution infeasible. However, it may be necessary to add many such additional constraints to find a sensible solution; this means that the solution is basically shaped by the constraints that the decision maker enforces in order to rule out blatantly inadequate portfolios. Incidentally, if short-selling is allowed, the decision variables are unrestricted, and the expected value of future wealth goes to infinity. In fact, one would short-sell assets with low expected return, to raise money to be invested in the most promising asset. This is clearly risky and should be carefully disciplined.

The next example is more akin to pricing a risky asset. It provides good evidence that pricing by the expected value of the payoff (possibly discounted, in order to take time value of money into account) does not seem a plausible approach.

Example 7.6 St. Petersburg paradox

Consider the following proposal. We are offered a lottery, whose outcome is determined by flipping a fair and memoryless coin. The coin is flipped until it lands tails. Let k be the number of times the coin lands heads; then, the payoff we get is \$$2^k$. Now, how much should we be willing to pay for this lottery? Even if we are unlucky and the game stops at the first flip, so that $k = 0$, we will get \$1, so we should be willing to pay at least this amount.

We may consider this as an asset pricing problem and set the expected value of the payoff as the fair price for this rather peculiar asset. The probability of winning \$$2^k$ is the probability of observing k consecutive heads followed by the tails that stops the game, after $k+1$ flips of the coin. Given the independence of events, the probability of this sequence is $1/2^{k+1}$, i.e., the product of $k + 1$ individual event probabilities. Then, the expected value of the payoff is

$$\begin{aligned}\sum_{k=0}^{\infty} \frac{1}{2^{k+1}}\, 2^k &= \tfrac{1}{2} \times 1 + \tfrac{1}{4} \times 2 + \tfrac{1}{8} \times 4 + \cdots \\ &= \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2} + \cdots \\ &= +\infty.\end{aligned}$$

This game looks so beautiful that we should be willing to pay any amount of money to play it! No one would probably do so. True, the game offers huge payoffs, but with vanishing probabilities. Again, we conclude that expected values do not tell the whole story.

The idea that most decision makers are risk-averse is intuitively clear, but what does **risk aversion** really mean in formal terms? To get a clue, let us compare two simple lotteries:

1. Lottery a_1, which is actually deterministic and guarantees a sure payoff μ
2. Lottery a_2, which offers two equally likely payoffs $\mu + \delta$ and $\mu - \delta$

The two lotteries are clearly equivalent in terms of expected payoff, but a risk-averse agent will arguably select lottery a_1. More generally, if we consider a random variable X, representing a payoff, and we add a **mean-preserving**

spread, i.e., an independent random variable $\tilde{\epsilon}$ with $\mathbb{E}[\tilde{\epsilon}] = 0$,[6] this addition is not welcome by a risk-averse decision maker and the lottery X is preferred to $X + \tilde{\epsilon}$. This idea may be further formalized and made operational by using different approaches that are discussed in the following.

7.3 A conceptual tool: The utility function

Given a set of lotteries, a decision maker should be able to pick the preferred one; or, given any pair of lotteries, the decision maker should be able to tell which one she prefers or state that she is indifferent between them. If so, she has a well-defined preference relationship among lotteries. Since preference relationships are a bit cumbersome and difficult to deal with, we could map each lottery to a real number measuring the attractiveness of that lottery to the decision maker, and then use the standard ordering of real numbers to rank lotteries. Such a function cannot be just the expectation, as this disregards risk aversion. A theoretical answer, commonly put forward in economic theory, can be found by assuming that decision makers order uncertain outcomes by a suitably chosen functional, rather than by straightforward expected monetary values. For an arbitrary preference relationship, a functional representing it may not exist but, under a set of more or less reasonable assumptions,[7] such a mapping does exist and can be represented by an **expected utility**. A particularly simple form of expected utility functional, which looks reasonable, but it is only justified by specific hypotheses on the preference relationship that it represents, is the **Von Neumann–Morgenstern expected utility**, defined as

$$U(X) = \mathbb{E}\big[u(X)\big],$$

for a suitably chosen function $u(\cdot)$. For a simple lottery a represented by a discrete random variable with n outcomes x_i and probabilities p_i, this boils down to

$$U(a) = \sum_{i=1}^{n} p_i u(x_i).$$

To be precise, we refer to function $u(\cdot)$ as the utility function, which is related to a certain payoff. On the contrary, $U(\cdot)$ is the expected utility *functional*, as it maps random variables to the real line. If $u(x) \equiv x$, then the expected utility functional boils down to the expected value of the payoff. Alternative choices of the utility function $u(\cdot)$ model different attitudes toward risk. For financial

[6] For the sake of convenience, when using Greek letters we denote by $\tilde{\epsilon}$ a random variable and by ϵ a realization of that variable. This notation is common in economics. In statistics, one typically uses X and x with the corresponding pair of meanings, but this is not quite convenient with Greek letters.

[7] The discussion of these assumptions is best left to books on microeconomics or decision theory; we should mention that most of them seem rather innocent and reasonable, under most circumstances, but they may lead to surprising effects in paradoxical cases.

problems, it is reasonable to assume that utility $u(\cdot)$ is a strictly increasing function, since we prefer more wealth to less. Formally, this property is referred to as **non-satiation**.

Beside the requirement of increasing monotonicity, the utility function is typically assumed to be concave. It is easy to see that concavity may express risk aversion. For the sake of convenience, we recall that a function f is said to be concave on a domain $S \subseteq \mathbb{R}^n$, if

$$f\big(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\big) \geq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in S, \lambda \in [0,1]. \tag{7.2}$$

In words, the value of the function for a convex combination of points in the domain is larger than the corresponding convex combination of the function values.[8] Since a convex combination is a linear combination with non-negative weights adding up to one, we immediately see the link with expected values. If we consider a lottery featuring two possible outcomes, x_1 and x_2, with probabilities $p_1 = p$ and $p_2 = 1 - p$, respectively, a risk-averse decision maker would prefer not taking chances:

$$u\big(\mathbb{E}[X]\big) = u\big(px_1 + (1-p)x_2\big) \geq pu(x_1) + (1-p)u(x_2) = \mathbb{E}\big[u(X)\big]. \tag{7.3}$$

This may be generalized to a generic, possibly continuous random variable by recalling **Jensen's inequality** for a concave function u of a random variable X:

$$u\big(\mathbb{E}[X]\big) \geq \mathbb{E}\big[u(X)\big]. \tag{7.4}$$

Example 7.7 Concavity and risk aversion

Let us consider again the sure lottery a_1, which guarantees a payoff μ with probability one, and lottery a_2, obtained by the mean-preserving spread $\tilde{\epsilon}$, featuring equally likely outcomes $-\delta$ and δ. Concavity implies risk aversion, since

$$U(a_1) = u(\mu) \geq \tfrac{1}{2}u(\mu - \delta) + \tfrac{1}{2}u(\mu + \delta) = U(a_2).$$

Since the inequality is not strict, we should say that lottery a_1 is at least as preferred as a_2, and the decision maker could be indifferent between the two.

As a numerical illustration, let us consider the logarithmic utility $u(x) = \log x$, and $\mu = 10$, $\delta = 5$:

$$U(a_1) = \log 10 = 2.3026,$$
$$U(a_2) = \tfrac{1}{2}\log 5 + \tfrac{1}{2}\log 15 = 2.1587.$$

Figure 7.1 illustrates the role of concavity in describing risk aversion.

[8] See Section 15.1 for more details on convex and concave functions.

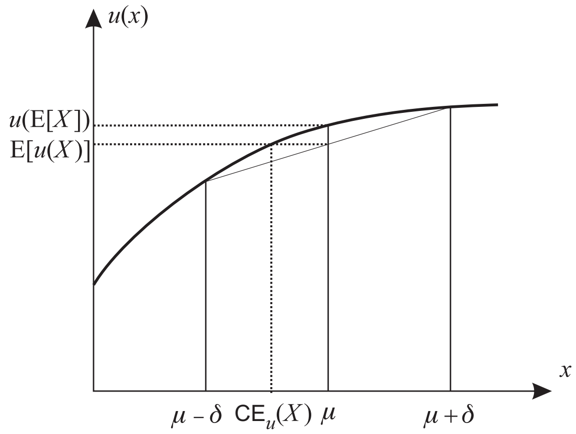

FIGURE 7.1 How concave utility functions imply risk aversion; the certainty equivalent is also shown.

It is fundamental to observe that the specific numerical value that the utility function assigns to a lottery is irrelevant per se; only the *relative* ordering of alternatives is essential. In fact, we speak of **ordinal** rather than *cardinal* utility. Given the linearity of expectation, we also see that an affine transformation of the utility function $u(\cdot)$ has no effect, provided it is increasing. To see this, let us consider $\bar{u}(x) \doteq au(x) + b$ instead of $u(x)$, where $a > 0$. Then, the ranking of alternatives according to u is clearly preserved by $\bar{u}$, since

$$\bar{U}(X) = \mathbb{E}\big[\bar{u}(X)\big] = \mathbb{E}\big[au(X) + b\big] = aU(X) + b.$$

Concavity implies risk aversion, from a qualitative viewpoint, but we would also like to come up with some *quantitative* way to measure risk aversion. We have said that a risk-averse decision maker would prefer a certain payoff to an uncertain one, when the expected values are the same. She would take the gamble only if the expected value of the risky lottery were suitably larger than the certain payoff. In other words, she requires a **risk premium**. The risk premium depends partly on the risk attitude of the decision maker, and partly on the uncertainty of the gamble itself. We will denote the risk premium by $\rho_u(X)$[9]; note that this is a number that a decision maker with utility $u(\cdot)$ associates with a random variable X. The risk premium is implicitly defined by the condition

$$u\big(\mathbb{E}[X] - \rho_u(X)\big) = U(X). \tag{7.5}$$

The risk premium also defines a **certainty equivalent**, i.e., a sure and guaranteed payoff $\mathsf{CE}_u(X)$, such that the agent would be indifferent between this certain amount and the uncertain lottery:

$$\mathsf{CE}_u(X) = \mathbb{E}[X] - \rho_u(X).$$

[9]Hopefully, no confusion will arise with the usual notation for the correlation coefficient ρ.

Note that the certainty equivalent is smaller than the expected value, and the difference is larger when the risk premium is larger. These concepts may be better grasped by looking again at Fig. 7.1.

Example 7.8 Certainty equivalent and risk premium

In Example 7.7, we have seen that the sure lottery a_1 is preferred to a_2 by a decision maker characterized by a logarithmic utility. Let us find the corresponding certainty equivalent for lottery a_2. We need a sure amount $x = \mathsf{CE}_{\log}(a_2)$, such that

$$u(x) = \log x = U(a_2) = 2.1587.$$

Hence,

$$\mathsf{CE}_{\log}(a_2) = e^{2.1587} = 8.6603,$$

and the risk premium is

$$\rho_{\log}(a_2) = 10 - 8.6603 = 1.3397.$$

We may interpret the risk premium as the additional expected payoff that a risk-averse decision maker requires to switch from the risk-free alternative a_1 to the risky alternative a_2, or the amount that she is willing to give up in order to get rid of the risk of a_2.

Example 7.8 points out a difficulty with the risk premium concept: It mixes the intrinsic risk of a lottery[10] with the subjective risk attitude of the decision maker. Thus, we might wish to separate the two sides of the coin. Consider a lottery $X = x + \tilde{\epsilon}$, where x is a given number and $\tilde{\epsilon}$ is a random variable with $\mathbb{E}[\tilde{\epsilon}] = 0$ and $\mathrm{Var}(\tilde{\epsilon}) = \sigma^2$. Hence,

$$\mathbb{E}[X] = x, \quad \mathrm{Var}(X) = \sigma^2.$$

Let us assume that the random variable $\tilde{\epsilon}$ is a “small” perturbation, in the sense that any possible realization ϵ is a relatively small number. Hence, we may approximate both sides of Eq. (7.5) by Taylor expansions. Consider, for instance, the expression $u(x+\epsilon)$. Since only numbers are involved here, we may write

$$u(x+\epsilon) \approx u(x) + \epsilon u'(x) + \tfrac{1}{2}\epsilon^2 u''(x).$$

By using this approximation for the random variable $\tilde{\epsilon}$, under the assumption that its realizations are small enough, and taking expected values, we may ap-

[10] Here we assume that the risk is related to objective probabilities, but the same concept would apply in the case of a subjective assessment of probabilities, if we disregard distributional ambiguity.

proximate the right-hand side of Eq. (7.5) as follows:

$$\begin{aligned} U(X) \doteq \mathbb{E}\big[u(x+\tilde{\epsilon})\big] &\approx \mathbb{E}\big[u(x)+\tilde{\epsilon}u'(x)+\tfrac{1}{2}\tilde{\epsilon}^2u''(x)\big] \\ &= u(x)+\mathbb{E}[\tilde{\epsilon}]u'(x)+\tfrac{1}{2}\mathbb{E}[\tilde{\epsilon}^2]u''(x) \\ &= u(x)+0\cdot u'(x)+\tfrac{1}{2}\mathrm{Var}(\tilde{\epsilon})u''(x) \\ &= u(x)+\tfrac{1}{2}\sigma^2u''(x). \end{aligned} \tag{7.6}$$

In the second-to-last line, we have used the well-known identity $\mathrm{Var}(\tilde{\epsilon}) = \mathbb{E}[\tilde{\epsilon}^2] - \mathbb{E}^2[\tilde{\epsilon}] = \mathbb{E}[\tilde{\epsilon}^2] - 0$. We may also approximate the left-hand side of Eq. (7.5), which involves only numbers, by a first-order expansion around $\mathbb{E}[X] = x$:

$$u\big(\mathbb{E}[X]-\rho_u(X)\big) \approx u(x)-\rho_u(X)u'(x). \tag{7.7}$$

By equating the two approximations (7.6) and (7.7) and rearranging, we find

$$\rho_u(X) = -\frac{1}{2}\frac{u''(x)}{u'(x)}\sigma^2. \tag{7.8}$$

Since we assume that the utility function is concave and strictly increasing, the right-hand side of Eq. (7.8) is well-defined and positive.[11] We observe that the risk premium is factored as the product of a term depending on the agent's subjective risk aversion, represented by the utility function $u(\cdot)$, and another one depending on the intrinsic uncertainty of the lottery, represented by the standard deviation σ. This justifies the following definition of the **coefficient of absolute risk aversion**:

$$R_u^a(x) \doteq -\frac{u''(x)}{u'(x)}. \tag{7.9}$$

The more concave the utility function, i.e., the larger $u''(x)$ in absolute value, the larger the risk aversion. We have observed that, given the linearity of the expectation operator, transforming the utility function $u(x)$ by an increasing affine transformation is inconsequential. Indeed, the definition of the risk aversion coefficient is consistent with this observation, as it is easy to see that the coefficients for $u(x)$ and $\bar{u}(x) = au(x) + b$ are the same.

We should also note that the coefficient $R_u^a(x)$ may change considerably as a function of x. If we consider the asset allocation problem of Example 7.3, we may use expected utility as the functional of terminal wealth $W_T(\omega)$, which we should maximize with respect to the vector $\mathbf{h}$ of the asset holdings. Let us denote by $R_{\mathbf{h}}$ the corresponding holding period return of the portfolio. Then, we should maximize

$$U(W_T) = \mathbb{E}\big[u\big(W_0\cdot(1+R_{\mathbf{h}})\big)\big].$$

In general, the solution may change as a function of W_0. From an investor's perspective, in fact, risk aversion may depend on the current level of wealth.

[11] We recall that, for a differentiable concave function of one variable, we have $u''(x) \leq 0$.

By a similar token, we may define the **coefficient of relative risk aversion**. This is motivated by considering a *multiplicative*, rather than additive, shock on an expected value x: $X = x \cdot (1 + \tilde{\epsilon})$. Here $\mathbb{E}[\tilde{\epsilon}] = 0$ and $\mathrm{Var}(\tilde{\epsilon}) = \sigma^2$, as before, but

$$\mathbb{E}[X] = x, \quad \mathrm{Var}(X) = x^2\sigma^2.$$

The mean is preserved again, but the random variable $\tilde{\epsilon}$ is related to a return in this case. Then, we may consider a *relative* risk premium $\pi_u(X)$ as the *fraction* of wealth that the decision maker is willing to give up in order to avoid taking chances,

$$\pi_u(X) \doteq \frac{x - \mathsf{CE}_u(X)}{x},$$

which implies

$$\rho_u(X) = x - \mathsf{CE}_u(X) = \pi_u(X) \cdot x.$$

Now, using a first-order Taylor approximation as before, we may write

$$\begin{aligned} u\big(\mathsf{CE}_u(X)\big) &= u\big(\mathbb{E}[X] - \rho_u(X)\big) \approx u(x) - \rho_u(X)u'(x) \\ &= u(x) - \pi_u(X) \cdot xu'(x). \end{aligned} \tag{7.10}$$

The utility of X can be approximated by a second-order expansion, for a realization ϵ:

$$u(X) = u(x + x\epsilon) \approx u(x) + u'(x)x\epsilon + \frac{1}{2}u''(x)x^2\epsilon^2.$$

By taking expectations and observing that $\mathbb{E}[\epsilon^2] = \sigma^2$, we find

$$\mathbb{E}\big[u(X)\big] \approx u(x) + \frac{1}{2}u''(x)x^2\sigma^2. \tag{7.11}$$

Putting Eqs. (7.10) and (7.11) together and rearranging yield

$$\pi_u(X) = -\frac{1}{2}\frac{u''(x)}{u'(x)}x\sigma^2,$$

which suggests the definition of the relative risk aversion coefficient,

$$R_u^r(x) \doteq -\frac{u''(x)x}{u'(x)}. \tag{7.12}$$

The only difference with respect to the absolute coefficient is the multiplication by x.

7.3.1 A FEW STANDARD UTILITY FUNCTIONS

Beside listing some common utility functions, in this section, we want to illustrate how to classify them according to some relevant criteria. This is best illustrated by a simple example.

Example 7.9 Logarithmic utility

A typical utility function is the logarithmic utility:

$$u(x) = \log(x). \tag{7.13}$$

Clearly this makes sense only for positive values of wealth. It is easy to check that, for the logarithmic utility, we have

$$R_u^a(x) = \frac{1}{x}, \qquad R_u^r(x) = 1.$$

Hence, logarithmic utility has decreasing absolute risk aversion, but constant relative risk aversion.

The coefficients of absolute and relative aversion may be decreasing, constant, or increasing with respect to their argument. Hence, utility functions may belong to one of the following families:

- Decreasing, or constant, or increasing absolute risk aversion, denoted by **DARA**, **CARA**, and **IARA**, respectively.
- Decreasing, or constant, or increasing relative risk aversion, denoted by **DRRA**, **CRRA**, and **IRRA**, respectively.

Thus, logarithmic utility is DARA and CRRA. Furthermore, it may be thought of as a limit case of the more general family of power utility functions:

$$u(x) = \frac{x^{1-\gamma} - 1}{1-\gamma}, \qquad \gamma > 1. \tag{7.14}$$

To understand the reasons behind the parameterization with respect to γ, let us find the coefficient of relative risk aversion of power utility:

$$\begin{aligned} u'(x) &= x^{-\gamma}, \\ u''(x) &= -\gamma x^{-(\gamma+1)}, \\ R_u^r(x) &= x \cdot \gamma \cdot \frac{x^{\gamma}}{x^{\gamma+1}} = \gamma. \end{aligned}$$

Furthermore, using L'Hôpital's rule,[12] we find[13]

$$\lim_{\gamma \to 1} \frac{x^{1-\gamma} - 1}{1-\gamma} = \lim_{\gamma \to 1} \frac{-\log(x) \cdot x^{1-\gamma}}{-1} = \log(x).$$

[12] L'Hôpital's rule is used to find the limit $\lim_{x \to x_0} f(x)/g(x)$, in the case where both functions $f(\cdot)$ and $g(\cdot)$ tend to zero. Subject to technical conditions, the limit is $\lim_{x \to x_0} f'(x)/g'(x)$.

[13] Here, we are also using the derivative of the function $f(x) = a^x = e^{x \log a}$, which is $f'(x) = a^x \cdot \log a$.

We may also consider the exponential utility function

$$u(x) = -e^{-\alpha x}, \tag{7.15}$$

for $\alpha > 0$. Note that this is an increasing function, and it is easy to interpret the parameter α:

$$R_u^a(x) = -\frac{-\alpha^2 e^{-\alpha x}}{\alpha e^{-\alpha x}} = \alpha.$$

Hence, we conclude that the exponential utility is CARA. This feature may be somewhat at odds with intuition, as one might expect that wealthier individuals are less averse to risk. It is important to remark that some utility functions have been used in the academic literature, because they are easy to manipulate, but this does not imply that they always model realistic investors' behavior.[14]

Another common utility function is **quadratic utility**:

$$u(x) = x - \frac{\lambda}{2}x^2. \tag{7.16}$$

Note that this function is not monotonically increasing and makes sense only for $x \in [0, 1/\lambda]$. Another odd property of quadratic utility is that it is IARA:

$$R_u^a(x) = \frac{\lambda}{1 - \lambda x} \quad \Rightarrow \quad \frac{dR_u^a(x)}{dx} = \frac{\lambda^2}{(1 - \lambda x)^2} > 0.$$

This implies, for instance, that an investor becomes more risk-averse if her wealth increases, which is usually considered at odds with standard investors' behavior. Nevertheless, it may be argued that, since any concave utility function may be locally approximated by a quadratic utility function, this provides a useful tool anyway. Furthermore, quadratic utility emphasizes the role of variance, since we have

$$U(X) = \mathbb{E}\left[X - \frac{\lambda}{2}X^2\right] = \mathbb{E}[X] - \frac{\lambda}{2}\left(\mathrm{Var}(X) + \mathbb{E}^2[X]\right). \tag{7.17}$$

A decision maker with quadratic utility is basically concerned only with the expected value and the variance of an uncertain outcome. In chapter 8, we will discuss the connection with mean–variance portfolio optimization (see Section 8.5).

Example 7.10 Logarithmic utility and portfolio choice

Consider the following stylized portfolio optimization problem:

- We represent uncertainty in asset return by a binomial model: There are two possible states of the world in the future, the up and down states, with probabilities p and $q = 1 - p$, respectively.

[14] See Problem 7.1 for an example concerning the odd behavior of the exponential utility function.

- There are two assets: one is risk-free, the other one is risky.
- The risk-free asset has gain R_f in both states (recall that multiplicative gain is one plus holding period return; in other words, \$1 grows to $\$R_f$).
- Current price for the risky asset is S_0 and its gain is u in the up state and d in the down state. Hence, the two possible risky asset prices are uS_0 and dS_0. We use gain, rather than holding period return, to streamline notation.
- Initial wealth is W_0 and the investor has logarithmic utility.

In this problem, there is actually one decision variable, which we may take as δ, the number of stock shares purchased by the investor. To get rid of the budget constraint, we observe that δS_0 is the wealth invested in the risky asset, and $W_0 - \delta S_0$ is invested in the risk-free asset. Then, future wealth will be, for each of the two possible states:

$$W_u = \delta S_0 u + (W_0 - \delta S_0)R_f = \delta S_0(u - R_f) + W_0 R_f,$$
$$W_d = \delta S_0 d + (W_0 - \delta S_0)R_f = \delta S_0(d - R_f) + W_0 R_f,$$

and expected utility is $p\log(W_u) + q\log(W_d)$. The problem is then

$$\max_{\delta} \quad p\log\left\{\delta S_0(u - R_f) + W_0 R_f\right\} + q\log\left\{\delta S_0(d - R_f) + W_0 R_f\right\}.$$

Let us write the first-order (stationarity) condition for optimality:

$$p \cdot \frac{S_0(u - R_f)}{\delta S_0(u - R_f) + W_0 R_f} + q \cdot \frac{S_0(d - R_f)}{\delta S_0(d - R_f) + W_0 R_f} = 0.$$

In order to solve for δ, we may rearrange the equation a bit:

$$\frac{\delta S_0(u - R_f) + W_0 R_f}{pS_0(u - R_f)} = -\frac{\delta S_0(d - R_f) + W_0 R_f}{qS_0(d - R_f)}.$$

Straightforward manipulations yield

$$\frac{\delta}{p} + \frac{W_0 R_f}{pS_0(u - R_f)} = -\frac{\delta}{q} - \frac{W_0 R_f}{qS_0(d - R_f)}$$

and

$$\delta\left[\frac{1}{p} + \frac{1}{q}\right] = -\frac{W_0 R_f\left[q(d - R_f) + p(u - R_f)\right]}{pqS_0(u - R_f)(d - R_f)}.$$

Then, one last step yields

$$\frac{\delta S_0}{W_0} = \frac{R_f\left[up + dq - R_f\right]}{(u - R_f)(R_f - d)}.$$

This relationship implies that the *fraction* of initial wealth invested in the risky asset does not depend on the initial wealth itself. We have derived this property in a simplified setting, but it holds more generally for logarithmic utility and is essentially due to its CRRA feature.

Example 7.10 shows how the features of each utility function may affect the solution of decision problems. One must be aware of the implied behavior, when choosing a specific utility function. Once again, we recall that, in this chapter, we are dealing with static decision problems. The definition of a utility function gets much more complicated in the case of multistage problems, as intertemporal issues arise.

7.3.2 LIMITATIONS OF UTILITY FUNCTIONS

Utility functions have been subjected to much criticism over the years:

- They rely on critical assumptions about the underlying preference relationships and may lead to paradoxes.
- They assume a significant degree of rationality in decision makers, who may be affected in real life by lack of information and cognitive limitations, leading to behavioral anomalies that are not explained within the standard utility framework. Some experiments shows that the observed behavior of decision makers may contradict the expected utility paradigm, as we discuss in Section 10.5.
- They aim at modeling subjective risk aversion, but a portfolio manager has to cope with multiple clients, and she should certainly *not* make decisions according to her own degree of risk aversion. Objective risk measures may be preferable.
- It is difficult to *elicit* a specific utility function from a decision maker.

In Section 7.4, we resort to an alternative approach, based on mean–risk models. The idea is to introduce an objective risk measure, which is a functional mapping random variables into real numbers, and trade expected profit/return against risk. This leads to a multiobjective optimization problem. As we have seen in Supplement S2.1, one possibility to cope with multiple objectives is to form a linear combination of two objective functions. For instance, when dealing with a random return R, a natural idea is to define a risk-adjusted expected return,

$$\mathbb{E}[R] - \tfrac{1}{2}\lambda \mathrm{Var}(R). \tag{7.18}$$

This mean–risk objective looks much like an expected quadratic utility, even though a comparison with Eq. (7.17) shows that they are not exactly the same. We shall introduce alternative risk measures to cope with asymmetric risks.

Before doing so, we may take advantage of the streamlined form of Eq. (7.18) to show how we might try to estimate the risk aversion coefficient λ in a simple case.[15]

Example 7.11 Estimating risk aversion

Say that we own a piece of real estate and we want to insure it against a disaster that may occur with probability p. If disaster strikes, our loss is 100% of the property value. Risk may be represented by a Bernoulli random variable:

- With probability p, return is -1 (we lose 100% of the property).
- With probability $1-p$, return is 0.

Then,

$$\mathbb{E}[R] = p \times (-1) + (1-p) \times 0 = -p,$$
$$\mathrm{Var}(R) = p \times (-1)^2 + (1-p) \times 0^2 - p^2 = p(1-p).$$

Note that the expected return is negative, as we are facing a potential loss. By abusing proper quadratic utility a little bit, let us consider the mean–risk form of Eq. (7.18),

$$U(R) = \mathbb{E}[R] - \tfrac{1}{2}\lambda \mathrm{Var}(R).$$

In this specific case, the utility score is, for a given risk aversion coefficient λ,

$$U = -p - \tfrac{1}{2}\lambda p(1-p).$$

We may consider insuring the property for a given premium. The more we are willing to pay, the more risk-averse we are. If we are willing to pay at most ν, then the utility of the certain equivalent loss of $-\nu$ is equal to the above utility score:

$$U = -\nu \quad \Rightarrow \quad \nu = p + \tfrac{1}{2}\lambda p(1-p).$$

As a reality check, observe that a risk-neutral investor ($\lambda = 0$) would just pay p, the expected loss. Given the insurance premium ν that we are willing to pay, this relationship allows to figure out a sensible value of λ, since

$$\lambda = \frac{2(\nu - p)}{p(1-p)}.$$

To get a more intuitive feeling, imagine that p is small ($1 - p \approx 1$), so that

$$\nu \approx p + \tfrac{1}{2}\lambda p.$$

[15] The example is borrowed from [3, Chapter 6].

Let us try a few values of λ:

$$\begin{aligned}
\lambda = 0 \quad &\Rightarrow \quad \nu = p, \\
\lambda = 1 \quad &\Rightarrow \quad \nu \approx 1.5p, \\
\lambda = 2 \quad &\Rightarrow \quad \nu \approx 2p, \\
\lambda = 3 \quad &\Rightarrow \quad \nu \approx 2.5p.
\end{aligned}$$

Therefore, for each unit increment in the risk aversion coefficient, we should be willing to pay another 50% of the expected loss.

In portfolio optimization, it is commonly agreed that λ ranges between 2 and 4.

7.4 Mean–risk models

The framework of expected utility suffers from the limitations that we have outlined in Section 7.3.2. Arguably, the most critical one is that a utility function mixes objective risk measurement and subjective risk aversion in decision-making. This is quite evident in the concept of risk premium. Hence, practitioners in financial industry prefer to rely on the concept of a risk measure. From a mathematical viewpoint, we should arguably talk of a risk *functional*, since what we need is a way to map a random variable $X(\omega)$, which is itself a function, to a real number:

$$\xi : X(\omega) \rightarrow \mathbb{R}.$$

We will use both terms interchangeably. Armed with a risk measure, we may tackle the problem of finding a satisfactory risk–reward tradeoff by using concepts of multiobjective optimization, as discussed in Section S2.1. This results in mean–risk optimization models.

If we choose variance or standard deviation as risk measures, we end up with the mean–variance portfolio optimization model that we have introduced in Section 2.1.1. Mean–variance optimization relies on variance for the sake of computational convenience, as this choice leads to a simple quadratic programming model. However, the underlying idea is actually using standard deviation as a risk measure. Standard deviation can be considered as a risk measure: the smaller, the better. However, while standard deviation captures the dispersion of a probability distribution, is it really a good risk measure? Example 7.2 clearly shows that symmetric risk measures, like standard deviation or variance, may fail with skewed distributions. As an alternative, we have considered value-at-risk, which is an asymmetric risk measure, in Section 2.2.2.

Value-at-risk is an example of asymmetric risk measure based on quantiles. However, we may easily define an asymmetric risk measure based on variance, namely, **semivariance**. If X is a random variable modeling profit or return, its semivariance is defined as

$$\mathbb{E}\Big[\big(\max\{0, \mu_X - X\}\big)^2\Big]. \tag{7.19}$$

In practice, we consider only negative deviations with respect to the expected value. The idea can be generalized and made more flexible, if we introduce negative deviations with respect to a minimum target that we wish to achieve, i.e., shortfall amounts. Let us denote the random terminal wealth associated with a portfolio by W_T. If we choose a target wealth $W_{\min}$, we may be interested in evaluating the portfolio performance in terms of **shortfall probability**,

$$\mathrm{P}\big\{W_T < W_{\min}\big\},$$

or **expected shortfall**,

$$\mathbb{E}\big[\max\{0, W_{\min} - W_T\}\big].$$

Shortfall is zero if we achieve or exceed the target, so we are penalizing underachievement in an asymmetric way. Expected shortfall, when used within portfolio optimization modeling, may result in simple linear programming problems.[16] To this aim, we should discretize the expectation by generating a finite set of scenarios, as customary in stochastic programming. If we wish to penalize large shortfalls more heavily, we may consider the expected squared shortfall,

$$\mathbb{E}\Big[\big(\max\{0, W_{\min} - W_T\}\big)^2\Big],$$

which may be tackled by quadratic programming.

How do these measures compare against each other? In order to provide a sensible answer, we must clarify the desirable properties of a risk measure.

7.4.1 COHERENT RISK MEASURES

A **single-period risk measure** is a functional $\xi(\cdot)$ mapping a random variable $X(\omega)$ to the real line. The random variable might be interpreted as the value of a portfolio, or a profit or loss, i.e., a change in value. Furthermore, loss might be relative with respect to an expected future target, or an absolute loss. In this section, we list some desirable properties of a risk measure. In the literature, different statements of these properties may be found, depending on the interpretation of $X(\omega)$. Here, we assume that the random variable represents a profit or the value of a portfolio. Hence, the larger the random variable, the better, but the risk measure is defined in such a way that it should be minimized.

The following set of properties characterizes a **coherent** risk measure:

[16] See Chapter 15.

- **Normalization.** Consider a random variable that is identically zero, $X \equiv 0$. It is reasonable to set $\xi(0) = 0$; if we do not hold any portfolio, we are not exposed to any risk.
- **Monotonicity**. If $X_1 \leq X_2$,[17] then $\xi(X_1) \geq \xi(X_2)$. In plain English, if the value of portfolio 1 is never larger than the value of portfolio 2, then portfolio 1 is at least as risky as portfolio 2.
- **Translation invariance**. If we add a fixed amount a to the portfolio, the risk measure is affected: $\xi(X + a) = \xi(X) - a$. If $a > 0$, risk is reduced.
- **Positive homogeneity**. Intuitively, if we double the amount invested in a portfolio, we double risk. Formally: $\xi(bX) = b\xi(X)$, for $b \geq 0$.
- **Subadditivity**. Diversification is expected to decrease risk; at the very least, diversification cannot increase risk. Hence, it makes sense to assume that the risk of the sum of two random variables should not exceed the sum of the respective risks: $\xi(X + Y) \leq \xi(X) + \xi(Y)$.

We are dealing only with a single-period problem; tackling multiperiod problems may complicate the matter further, introducing issues related to time consistency, which we do not consider in this book.[18]

Remark. An interesting implication of translation invariance is

$$\xi\big(X + \xi(X)\big) = \xi(X) - \xi(X) = 0.$$

Thus, the risk measure of a portfolio with random value X may be interpreted as the minimum amount of additional capital that is needed to make the portfolio *acceptable*, where a portfolio X is said to be acceptable if its risk measure is $\xi(X) \leq 0$. In fact, risk measures (functionals) may also be interpreted as **acceptability functionals**.

We have listed theoretical requirements of a risk measure, but what about the practical ones?

- Clearly, a risk measure should not be overly difficult to compute. Unfortunately, computational effort may be an issue, if we deal with financial derivatives whose pricing itself requires intensive computation.
- When solving a portfolio optimization model, convexity is a quite important feature. Positive homogeneity and subadditivity may be combined into a convexity condition:

  $$\xi\big(\lambda X + (1 - \lambda)Y\big) \leq \lambda\xi(X) + (1 - \lambda)\xi(Y), \qquad \forall \lambda \in [0, 1].$$

  Thus, apart from theoretical considerations, a coherent risk measure may be practically preferable from a computational viewpoint.

[17] Since we are comparing random variables, the inequality should be qualified as holding almost surely, i.e., for all of the possible outcomes, with the exception of a set of measure zero. The unfamiliar reader may consider this as a technicality.

[18] The essence of time consistency of a multiperiod risk measure is that if a portfolio is riskier than another portfolio at time horizon τ, then it is riskier at time horizons $t < \tau$ as well. See, e.g., [2].

- Another requirement is that the risk measure should be easily communicated to top management. A statistically motivated measure, characterizing a feature of a probability distribution, may be fine for the initiated, but a risk measure expressed in hard monetary terms can be easier to grasp. We also note that specific sensitivity measures, like bond duration (and the option Greeks that we shall meet later), do not enable us to summarize all risk contributions, irrespectively of the nature of the different positions held in the portfolio. These difficulties led to the development of value-at-risk.

7.4.2 STANDARD DEVIATION AND VARIANCE AS RISK MEASURES

We are aware that a major limitation of standard deviation and variance is their symmetry, since they measure dispersion without paying attention to direction of variability. Let us run a more formal check by asking whether they meet the coherence requirements.

- The normalization requirement is met, but we know that, for any real number a,
$$\mathrm{Var}(X + a) = \mathrm{Var}(X).$$
Hence, variance and standard deviation are not translation invariant. We find a translation invariant measure, however, if we consider
$$\xi(X) = -\mathbb{E}[X] + \lambda\sqrt{\mathrm{Var}(X)},$$
since
$$\xi(X + a) = -\mathbb{E}[X + a] + \lambda\sqrt{\mathrm{Var}(X + a)} = \xi(X) - a.$$
- Monotonicity fails, as we have seen in Example 7.2. More generally, if we have a random variable bounded by a constant,
$$X_1(\omega) \le \alpha,$$
and we consider $X_2(\omega) = \alpha$, the monotonicity condition fails since $\mathrm{Var}(X_1) \ge 0$ and $\mathrm{Var}(X_2) = 0$.
- Let us consider positive homogeneity. Since
$$\mathrm{Var}(bX) = b^2\mathrm{Var}(X),$$
this condition fails for variance, but it is met by standard deviation (just take the square root).
- Let us complete the picture with subadditivity. Since
$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2),$$

variance fails to meet subadditivity when covariance is positive. However, if we consider standard deviation and we express covariance using the correlation coefficient $\rho_{12} \leq 1$, we see that standard deviation is subadditive:

$$\begin{aligned}\sigma_{X_1+X_2} &= \sqrt{\sigma_{X_1}^2 + \sigma_{X_2}^2 + 2\rho_{12}\sigma_{X_1}\sigma_{X_2}} \\ &\leq \sqrt{\sigma_{X_1}^2 + \sigma_{X_2}^2 + 2\sigma_{X_1}\sigma_{X_2}} = \sigma_{X_1} + \sigma_{X_2}.\end{aligned}$$

Hence, the picture is not quite encouraging for standard deviation and variance as risk measures, but standard deviation looks a bit better. From a practical viewpoint, when dealing with the return of a simple portfolio, variance may result in simple optimization problems, i.e., convex quadratic programs. However, this is not necessarily true when considering more complicated optimization models, where scenarios in terms of underlying risk factors are generated and mapped to asset prices by a nonlinear pricing model, and a stochastic programming model is solved.[19] Furthermore, while standard deviation or return or wealth may make sense, variance of wealth, which is measured in squared monetary units, cannot be really be interpreted. Even standard deviation of wealth may fail to convey a precise perception of directional risk. Nevertheless, these measures are broadly used in the context of modern portfolio theory, which relies on mean–variance optimization. As we shall see, this provides us with useful insights, like the capital asset pricing model, and it may be sometimes justified, since quadratic utility can approximate a generic concave utility function locally.

7.4.3 QUANTILE-BASED RISK MEASURES: V@R AND CV@R

We have introduced value-at-risk, in Section 2.2.2.1, as a quantile of the probability distribution of loss. There, we have considered typical textbook examples relying on normality, in order give a simple picture. However, in practice, estimating V@R is far from trivial for a complex trading book involving exotic derivatives, as well as equity or fixed-income assets. Whatever approach we use for its computation, V@R is not free from some fundamental flaws, which depend on its definition as a quantile. We should be well aware of them, especially when using sophisticated computational tools that may lure us into a false sense of security. The following example shows how a quantile cannot distinguish between different tail shapes.

Example 7.12 Different shapes of a tail

Consider the two loss densities in Fig. 7.2. In Fig. 7.2(a), we observe a normally distributed loss and its 95% V@R, which is just its quantile

[19] See Chapter 15.

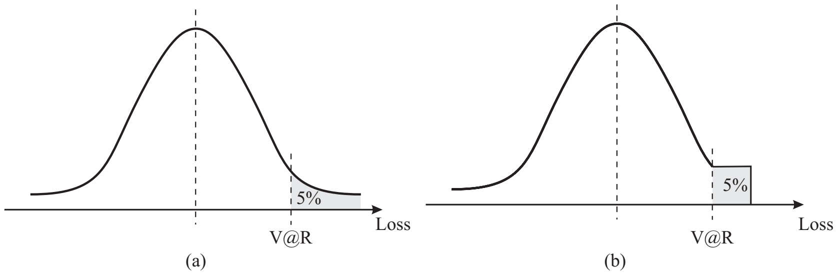

FIGURE 7.2 **Value-at-risk can be the same in quite different situations.**

at probability level 95%; the area of the right tail is 5%. In Fig. 7.2(b), we observe a sort of truncated distribution, obtained by replacing the tail of the normal PDF with a uniform density. The tail accounts for 5% of the total probability. By construction, V@R is the same in both cases, since the areas of the right tails are identical. However, we might not associate the same risk with the two distributions. In the case of the normal distribution, there is no upper bound to loss; in the second case, there is a clearly defined worst-case loss. Whether the risk for density (a) is larger than density (b) or not, it depends on how we measure risk exactly; the point is that V@R cannot tell the difference between them.

One way to overcome the limitations of a straightforward quantile, while retaining some of its desirable features, is to resort to a conditional expectation on the tail. This observation has led to the definition of alternative risk measures, such as **conditional value-at-risk** (CV@R), which (informally) is the expected value of loss, conditional on being to the right of V@R. For instance, the conditional (tail) expectation yields the midpoint of the uniform tail in the truncated density of Fig. 7.2(b); the tail expectation may be larger in the normal case of Fig. 7.2(a), because of its unbounded support. In this section, we investigate the properties of both V@R and CV@R in terms of coherence and computational viability.

7.4.3.1 A remark on quantiles

When defining quantile-based risk measures, there is no particular difficulty with standard continuous distributions featuring a continuous and strictly increasing CDF

$$F_X(x) \doteq \mathrm{P}\{X \leq x\}.$$

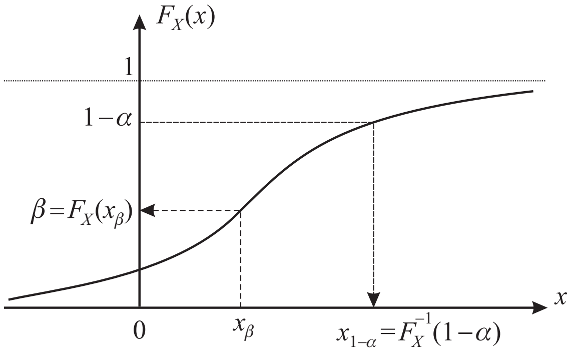

FIGURE 7.3 **The link between quantiles and the CDF.**

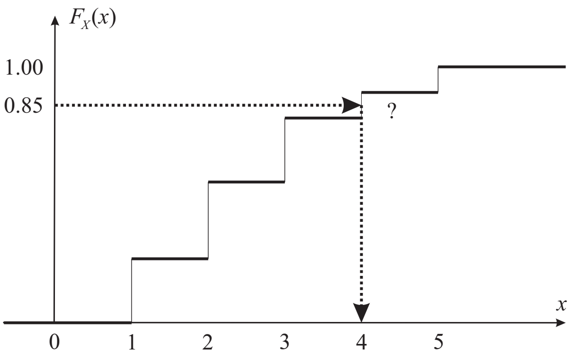

FIGURE 7.4 **A noninvertible CDF of a discrete random variable.**

In such a case, the CDF is invertible and the quantile $x_{1-\alpha}$ at probability level $1-\alpha$ is easily found:

$$F_X(x_{1-\alpha}) = 1-\alpha \quad \Rightarrow \quad x_{1-\alpha} = F_X^{-1}(1-\alpha), \tag{7.20}$$

where $\alpha \in [0,1]$ is the probability mass on the right tail (which is supposed to be small if we are considering a loss). This is illustrated in Fig. 7.3. Given a numerical value x_β, the CDF $F_X(x_\beta)$ gives the corresponding probability $\beta = \mathrm{P}\{X \le x_\beta\}$. Going the other way around, given the probability $1-\alpha$, inversion of the CDF yields the corresponding quantile $x_{1-\alpha}$.

However, the case of a discrete random variable is more involved, as the CDF is piecewise constant and not invertible. Figure 7.4 shows the CDF for a discrete random variable with the following PMF:

x	0	1	2	3	4	5
$p_X(x)$	0.05	0.20	0.30	0.25	0.10	0.10

Here,

$$F_X(3) = 0.8, \qquad F_X(4) = 0.9,$$

and we are in trouble when looking for the quantile $x_{0.85}$. Then, we may define the quantile as the smallest number $x_{1-\alpha}$ such that $F_X(x_{1-\alpha}) \geq 1-\alpha$. This relies on the definition of a **generalized inverse function**:

$$x_{1-\alpha} = \min\{x : F_X(x) \geq 1-\alpha\}.$$

The generalized inverse function boils down to the standard inverse, when the CDF is continuous and strictly increasing. In the numerical case that we are considering, we find

$$x_{0.85} = 4,$$

which makes sense in terms of "staying on the safe side." The intuitive idea is that the value 4 "covers" loss with a 90% guarantee, which is larger than necessary, but the value 3 offers a guarantee of only 80%. This looks innocent enough, but we should wonder about the possibility of defining quantiles using a *strict* inequality, as in

$$\inf\{x : F_X(x) > 1-\alpha\},$$

where we really have to use $\inf$, since the strict inequality does not guarantee the existence of a minimum. Actually, the question is not trivial and leads to alternative definitions of quantiles and risk measures, which may differ in terms of coherence. For the sake of simplicity, we will cut a few corners as usual.[20]

7.4.3.2 Is value-at-risk coherent?

In this section, we consider a continuous random variable L_T, modeling loss over the time horizon T for which we want to evaluate V@R. We assume that its CDF is invertible, so that value-at-risk with confidence level $1-\alpha$, for the given time horizon T, is the usual quantile $\mathrm{V@R}_{1-\alpha,T}$, such that

$$\mathrm{P}\{L_T \leq \mathrm{V@R}_{1-\alpha,T}\} = 1-\alpha, \tag{7.21}$$

and we may disregard technical complications. If we consider an affine transformation $aL_T + b$ of loss, with $a \geq 0$, we may manipulate Eq. (7.21) and find

$$\mathrm{P}\{aL_T + b \leq a\mathrm{V@R}_{1-\alpha,T} + b\} = 1-\alpha,$$

showing translation invariance[21] and positive homogeneity. Value-at-risk is clearly normalized and monotonic, but what about subadditivity? If we restrict our attention to specific classes of distributions, such as the normal, V@R is subadditive (see Problem 7.4). However, this depends on the fact that quantiles of a normal distribution are related to the standard deviation, which is subadditive. The following counterexample is often used to show that V@R is not subadditive in general.

[20] For a deeper analysis, see, e.g., [6].

[21] In this case, the constant b is added, rather than subtracted, which seems at odds with the previous definition of translation invariance. The point is that the random variable L_T represents a loss, rather than a profit.

Example 7.13 V@R is not subadditive

Let us consider two zero-coupon bonds, whose issuers may default with probability 4% (over some time horizon that we leave implicit). Say that, in the case of default, we lose the full face value, \$100 (in practice, we might partially recover the face value of the bond). Let us compute the V@R of each bond with confidence level 95%. We represent the loss for the two bonds by random variables X and Y, respectively, which take values in the set $\{0, 100\}$. Since loss has a discrete distribution in this example, we should use the more general definition of V@R provided by the generalized inverse. The probability of default is 4%, and $1 - 0.04 = 0.96 > 0.95$; therefore, we find

$$\begin{aligned} &\mathrm{V@R}_{0.95}(X) = \mathrm{V@R}_{0.95}(Y) = \$0 \\ &\quad \Rightarrow \mathrm{V@R}_{0.95}(X) + \mathrm{V@R}_{0.95}(Y) = \$0. \end{aligned}$$

Now what happens if we hold both bonds and assume independent defaults? We will suffer:

- A loss of \$0, with probability $0.96^2 = 0.9216$
- A loss of \$100, with probability $2 \times 0.96 \times 0.04 = 0.0768$
- A loss of \$200, with probability $0.04^2 = 0.0016$

Now the probability of losing \$0 is smaller than 95%, and

$$\mathrm{P}\{X + Y \leq 100\} = 0.9216 + 0.0768 > 0.95.$$

Hence, with that confidence level,

$$\mathrm{V@R}_{0.95}(X + Y) = 100 > \mathrm{V@R}_{0.95}(X) + \mathrm{V@R}_{0.95}(Y),$$

which means that risk, as measured by V@R, may be increased by diversification.

The lack of subadditivity also implies that minimization of V@R may not result in convex portfolio optimization problems. When uncertainty is represented by sampled scenarios, it turns out that it is not even a differentiable function of portfolio weights.[22] This is not to say that V@R is not useful and relevant. In fact, it is used to assess capital requirements for banks, i.e., to determine the liquidity needed as a buffer against short-term loss. However, it must be used with care, and alternative measures have been proposed for the same purpose.

[22] See, e.g., [5, pp. 615–618].

7.4.3.3 Conditional value-at-risk

Conditional value-at-risk, CV@R, is an asymmetric risk measure related to tail expectations and, as such, bears some similarity with expected shortfall (in fact, the two concepts are sometimes confused). However, in expected shortfall, we fix a target a priori; here, the threshold is given by V@R. Informally, CV@R is defined as a conditional tail expectation[23] of loss over a time horizon T, where the threshold is V@R with probability level $1-\alpha$:

$$\mathrm{CV@R}_{1-\alpha,T} \doteq \mathbb{E}\big[L_T \mid L_T > \mathrm{V@R}_{1-\alpha,T}\big]. \tag{7.22}$$

Since CV@R looks like a complication of V@R, it seems reasonable to expect that it is an even more difficult beast to tame. On the contrary, CV@R is much better behaved:

- It can be shown that CV@R is a coherent risk measure.
- A consequence of coherence is that CV@R is a convex risk measure.

The last point is quite relevant in terms of optimization modeling, as it suggests that minimization of CV@R and optimization subject to an upper bound on CV@R may result in relatively simple convex problems. We will consider CV@R optimization later, in Section 15.6.2.1. For now, let us consider an example in which, rather unsurprisingly, CV@R is easy to find.

Example 7.14 CV@R **in the normal case**

In the case of a normally distributed loss, $L \sim \mathsf{N}(\mu_L, \sigma_L^2)$, we may find an explicit expression for CV@R. Let us consider a standard normal loss $Z \sim \mathsf{N}(0,1)$ first, where $\mathrm{V@R}_{1-\alpha} = z_{1-\alpha}$, the familiar quantile for the standard normal distribution. We have

$$\begin{aligned}
&\mathbb{E}[Z \,|\, Z > z_{1-\alpha}] \\
&\quad = \frac{1}{\alpha}\int_{z_{1-\alpha}}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\
&\quad = \frac{1}{\alpha}\frac{1}{\sqrt{2\pi}}\int_{z_{1-\alpha}^2/2}^{+\infty} e^{-y} dy \quad \text{(variable substitution } y = x^2/2) \\
&\quad = \frac{1}{\alpha}\frac{1}{\sqrt{2\pi}} e^{-y}\Big|_{+\infty}^{z_{1-\alpha}^2}
\end{aligned}$$

[23] As we have already noted, there are some critical issues in the careful definition of quantile-based risk measures, especially when dealing with discrete distributions. We disregard such subtleties. We should also mention that the term "average value-at-risk" is also used to refer to CV@R.

$$= \frac{1}{\alpha}\frac{1}{\sqrt{2\pi}}e^{-z_{1-\alpha}^2/2} = \frac{1}{\alpha}\phi(z_{1-\alpha}),$$

where $\phi(z)$ is the PDF of the standard normal, as usual. For instance, if $\alpha = 0.05$ (note that this is the small area on the right tail of the loss distribution),

$$\mathbb{E}[Z \,|\, Z > z_{0.95}] = \frac{1}{0.05} \times \phi(1.6449) = 2.0627.$$

In the case of a generic normal loss $L \sim \mathsf{N}(\mu, \sigma^2)$, we just destandardize by considering

$$L = \mu + \sigma Z, \quad q_{1-\alpha} = \mu + \sigma z_{1-\alpha}.$$

Hence,

$$\begin{aligned} \mathbb{E}[L \,|\, L > q_{1-\alpha}] &= \mathbb{E}[\mu + \sigma Z \,|\, Z > z_{1-\alpha}] \\ &= \mu + \sigma \cdot \mathbb{E}[Z \,|\, Z > z_{1-\alpha}] \\ &= \mu + \frac{\sigma}{\alpha} \cdot \phi(z_{1-\alpha}). \end{aligned} \tag{7.23}$$

For instance, if $L \sim \mathsf{N}(-50, 200^2)$, where the negative expected loss corresponds to a positive expected profit of 50, we find

$$\mathrm{CV@R}_{0.95} = -50 + \frac{200}{0.05} \times \phi(1.6449) = 362.54.$$

7.4.4 FORMULATION OF MEAN–RISK MODELS

The exact formulation of a mean–risk model depends on both modeling and computational convenience. As we have seen in Supplement S2.1, there are different scalarization strategies to boil a multiobjective problem down to a sequence of single-objective problems. In the mean–risk case, we are dealing with a model where:

- We represent a portfolio by the vector $\mathbf{x}$ of decision variables, constrained by a feasible set $S \subset \mathbb{R}^n$.
- We want to maximize an expected profit/return $\pi(\mathbf{x})$.
- We want to minimize a risk measure $\xi(\mathbf{x})$.

In practice, we must select a scalarized model, which may be obtained by defining a risk-adjusted mean,

$$\begin{aligned} \max \quad & \pi(\mathbf{x}) - \lambda\xi(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in S, \end{aligned}$$

or by requiring a minimum expected reward,

$$\begin{aligned} \min \quad & \xi(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in S, \\ & \pi(\mathbf{x}) \geq \beta, \end{aligned}$$

or by defining a risk budget,

$$\begin{aligned} \max \quad & \pi(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in S, \\ & \xi(\mathbf{x}) \leq \gamma. \end{aligned}$$

Depending on the selected risk measure and the adopted scalarization, we formulate one of the mathematical programming problems to be discussed later, in Chapter 15. They range from manageable linear programming models to difficult nonconvex problems. Furthermore, the scalarizations involve a choice of parameters λ, β, or γ, which may have a more or less intuitive meaning to the decision maker. Efficient solvers are available for a wide class of (convex) optimization problems, enabling us to tackle many practically significant problems.

7.5 Stochastic dominance

In principle, the framework of utility functions allows to find a complete ordering of portfolios. However, utility functions are difficult to elicit, and an investor might be reluctant to commit to a specific utility. The mean–risk framework may provide us with a partial ordering of alternatives, as well as a set of efficient portfolios. The stochastic dominance framework is a third alternative framework, resulting in a partial ordering that may be related to broad families of utility functions.

To get the intuition and a possible motivation, let us consider again the two lotteries of Example 7.2. The example shows a limitation of mean–variance analysis, since one lottery is clearly dominated by the other one, yet, we have an unclear tradeoff in terms of mean and variance. We may introduce a concept of dominance between random returns/payoffs X and Y fairly easily. We say that X dominates Y if[24]

$$Y(\omega) \leq X(\omega), \quad \forall \omega \in \Omega, \tag{7.24}$$

[24] As usual, the condition should be better qualified, as it applies with the possible exception of a subset of the sample space Ω with null measure. If the random variables are discrete, this is not relevant.

and

$$P\{Y < X\} > 0.$$

In other words, X is never worse than Y, and X is strictly better than Y in some scenarios. Note that there is no clear relationship between this concept of dominance and efficiency. Nevertheless, assuming that investors are nonsatiated, i.e., they prefer more to less, no one would prefer Y to X. This concept of (strict) dominance is quite simple and intuitive, but it is not likely to be very useful in practice. It is unlikely that it will establish a rich preference relationship between portfolios. Actually, under a no-arbitrage assumption, we should expect that we *never* detect this kind of dominance.[25]

Hence, we must weaken the idea of strict dominance in order to find a more useful concept. To get a further clue, let us fix a target payoff/return β and assume that

$$P\{X \le \beta\} \;\le\; P\{Y \le \beta\}.$$

What does this condition suggest about the choice between X and Y? Since $P\{X \le \beta\} = 1 - P\{X > \beta\}$, we may rephrase the condition in terms of complementary probabilities as follows:

$$P\{X > \beta\} \;\ge\; P\{Y > \beta\}.$$

If we consider β as a target performance, we see that the probability of exceeding the target is larger for X than for Y. This may suggest that X is a better investment than Y, but actually this conclusion is not warranted, as the relationship could be reversed for other values of the target β. However, if we assume that this relationship holds for *every* possible target, we come up with the following definition.

DEFINITION 7.1 (First-order stochastic dominance) *Consider random variables X and Y. We say that X has first-order stochastic dominance over Y if*

$$P\{X \le \beta\} \;\le\; P\{Y \le \beta\}, \qquad \forall \beta \in \mathbb{R}$$

and

$$P\{X \le \gamma\} \;<\; P\{Y \le \gamma\}, \qquad \textit{for some } \gamma \in \mathbb{R}.$$

Note that the condition in Definition 7.1 may be restated in terms of the CDF of the two random variables:

$$\begin{aligned} F_X(\beta) &\le F_Y(\beta), && \forall \beta \in \mathbb{R} \\ F_X(\gamma) &< F_Y(\gamma), && \text{for some } \gamma \in \mathbb{R}. \end{aligned}$$

In plain English, if we plot the two CDFs, F_X is never above F_Y, and it is strictly less somewhere.

[25] See Section 2.4 for the link between dominance and arbitrage opportunities.

Table 7.3 **An example of first-order stochastic dominance.**

State	ω_1	ω_2	ω_3	ω_4	ω_5
Probability	0.2	0.2	0.2	0.2	0.2
Return r_X (%)	3	4	5	6	7
Return r_Y (%)	7	6	5	3	3

Return β (%)	2	3	4	5	6	7	8
$\mathrm{P}\{r_X \leq \beta\}$	0.0	0.2	0.4	0.6	0.8	1.0	1.0
$\mathrm{P}\{r_Y \leq \beta\}$	0.0	0.4	0.4	0.6	0.8	1.0	1.0

Example 7.15 First-order stochastic dominance

Let us consider the two investments described in Table 7.3. The first table gives the percentage return of the two investments in five states of the world. Clearly, there is no state-by-state dominance between the returns of the two alternatives. The second table shows the two CDFs for relevant values of return. Note that the CDF does not bear any relationship with the states of the world. The CDF of r_X never exceeds the CDF of r_Y and is strictly less at one point (return 3%). Hence, r_X first-order stochastically dominates r_Y.

In Definition 7.1, we are essentially assuming that investors prefer more to less, which is expressed by a strictly increasing utility function. This fact is formalized as follows.

THEOREM 7.2 *If X and Y satisfy the condition in Definition 7.1, then*

$$\mathbb{E}\big[u(X)\big] > \mathbb{E}\big[u(Y)\big],$$

for every utility function u satisfying the condition $u'(x) > 0$ for all x (u is differentiable and strictly increasing).

Actually, it turns out that the condition is necessary and sufficient, and we shall just sketch a proof in Supplement S7.1. We may get a glimpse of intuition by considering the following relationship *in distribution* between random variables X and Y:

$$Y \stackrel{d}{=} X + \xi, \tag{7.25}$$

where ξ is a nonpositive random variable. We should carefully note the fundamental difference between Eq. (7.24) and Eq. (7.25). In the latter case, we are *not* requiring a strong state-by-state condition,

$$Y(\omega) = X(\omega) + \xi(\omega), \qquad \forall \omega \in \Omega,$$

with $\xi(\omega) \leq 0$, but only a weaker condition in terms of distribution, which is actually a way to rephrase first-order stochastic dominance. Then, if the utility function u is strictly increasing, we find

$$\mathbb{E}\big[u(Y)\big] = \mathbb{E}\big[u(X+\xi)\big] < \mathbb{E}\big[u(X)\big].$$

First-order stochastic dominance is easier to observe in the real world than an unreasonable state-by-state dominance, but it is still too strong and may not allow to compare alternatives in many significant cases. To see why, let us consider the specific case $u(x) = x$, i.e., the utility function is the identity function, which is to say that the investor prefers more to less but is risk-neutral. We clearly see that the condition in Theorem 7.2 implies

$$\mathbb{E}[X] > \mathbb{E}[Y].$$

This means that we cannot compare distributions with the same expected value, which is a significant limitation. To overcome this difficulty, a weaker condition has been introduced.

DEFINITION 7.3 (Second-order stochastic dominance) *Let us consider random variables X and Y. We say that X has second-order stochastic dominance over Y if*

$$\int_{-\infty}^{\beta} \mathrm{P}\{X \leq s\}\, ds \leq \int_{-\infty}^{\beta} \mathrm{P}\{Y \leq s\}\, ds, \qquad \forall \beta \in \mathbb{R},$$

and

$$\int_{-\infty}^{\gamma} \mathrm{P}\{X \leq s\}\, ds < \int_{-\infty}^{\gamma} \mathrm{P}\{Y \leq s\}\, ds, \qquad \textit{for some } \gamma \in \mathbb{R}.$$

Definition 7.3 involves integrals of the CDF of random variables, which we may denote by

$$\widetilde{F}_X(x) \doteq \int_{-\infty}^{x} F_X(s)\, ds \equiv \int_{-\infty}^{x} \mathrm{P}\{X \leq s\}\, ds. \tag{7.26}$$

Hence, the condition of second-order stochastic dominance may be restated as follows:

$$\begin{aligned} &\widetilde{F}_X(\beta) \leq \widetilde{F}_Y(\beta), \qquad && \forall \beta \in \mathbb{R} \\ &\widetilde{F}_X(\gamma) < \widetilde{F}_Y(\gamma), && \text{for some } \gamma \in \mathbb{R}. \end{aligned}$$

First-order stochastic dominance implies second-order dominance; hence, it is a stronger concept. This is reflected in a weakened version of Theorem 7.2, whereby we add a condition related to risk aversion.

THEOREM 7.4 *If X and Y satisfy the condition in Definition 7.3, then*

$$\mathbb{E}\big[u(X)\big] > \mathbb{E}\big[u(Y)\big],$$

for every utility function u satisfying the conditions $u'(x) > 0$ and $u''(x) < 0$ for all x (u is differentiable, strictly increasing, and concave).

Stochastic dominance is an interesting concept, allowing us to establish a partial ordering between portfolios, which applies to a large range of sensible utility functions. Unfortunately, it is not quite trivial to translate the concept into computational terms, in order to make it suitable to portfolio optimization. Nevertheless, it is possible to build optimization models including stochastic dominance constraints with respect to a benchmark portfolio (see the chapter references).

S7.1 Theorem proofs

S7.1.1 PROOF OF THEOREM 7.2

The proof that we sketch here is rather limited, as we only deal with the case of random variables with a common bounded support $[a, b]$, for finite $a, b \in \mathbb{R}$. Nevertheless, it is simple enough and rather instructive. We assume that the random variables X and Y are continuous with densities (PDFs) $f_X(x)$ and $f_Y(y)$, related with the CDF as usual:

$$F_X(x) \doteq \mathrm{P}\{X \le x\} \qquad \text{and} \qquad f_X(x) = F_X'(x).$$

We assume differentiability throughout. We should consider the difference of the expected utilities, which may be written as follows:

$$\begin{aligned}\mathbb{E}[u(X)] - \mathbb{E}[u(Y)] &= \int_a^b u(x) f_X(x)\,dx - \int_a^b u(y) f_Y(y)\,dy \\ &= \int_a^b u(x) F_X'(x)\,dx - \int_a^b u(y) F_Y'(y)\,dy.\end{aligned}$$

Now we use integration by parts for both integrals. For instance,

$$\begin{aligned}\int_a^b u(x) F_X'(x)\,dx &= u(b)F_X(b) - u(a)F_X(a) - \int_a^b u'(x) F_X(x)\,dx \\ &= u(b) - \int_a^b u'(x) F_X(x)\,dx,\end{aligned}$$

since the assumption of bounded support implies $F_X(b) = 1$ and $F_X(a) = 0$. A similar relationship applies to Y, and we find

$$\begin{aligned}\mathbb{E}[u(X)] - \mathbb{E}[u(Y)] &= \int_a^b u'(y) F_Y(y)\,dy - \int_a^b u'(x) F_X(x)\,dx \\ &= \int_a^b u'(z)\Big[F_Y(z) - F_X(z)\Big]\,dz. \qquad (7.27)\end{aligned}$$

Now we observe that, for every z, $u'(z) > 0$, by the assumption of increasing monotonicity, and $F_Y(z) - F_X(z) \ge 0$, by the assumption of first-order dominance. Hence, the integral is positive, which proves the theorem.

S7.1.2 PROOF OF THEOREM 7.4

The proof, under similar assumptions about bounded support and differentiability of the involved functions, is quite similar to that of Theorem 7.2. We use the function $\widetilde{F}_X(x)$, i.e., the integral of the CDF that we have introduced in Eq. (7.26), so that

$$F_X(x) = \widetilde{F}'_X(x).$$

We start from Eq. (7.27) and, since risk aversion involves the second-order derivative $u''(x)$, we integrate by parts once more as follows:

$$\begin{aligned}
\mathbb{E}[u(X)] - \mathbb{E}[u(Y)] &= \int_a^b u'(z)\Big[F_Y(z) - F_X(z)\Big]\,dz \\
&= u'(b)\Big[\widetilde{F}_Y(b) - \widetilde{F}_X(b)\Big] - \underbrace{u'(a)\Big[\widetilde{F}_Y(a) - \widetilde{F}_X(a)\Big]}_{=0} \\
&\quad - \int_a^b u''(z)\Big[\widetilde{F}_Y(z) - \widetilde{F}_X(z)\Big]\,dz,
\end{aligned}$$

where the second term vanishes, since $\widetilde{F}_X(a)$ and $\widetilde{F}_X(a)$ are integrals on an interval $[a, a]$ with zero measure. Hence,

$$\widetilde{F}_Y(a) = \widetilde{F}_X(a) = 0.$$

The result now follows from the assumptions $u'(z) > 0$, $u''(z) < 0$, and the definition of second-order stochastic dominance.

Problems

7.1 Consider an exponential utility function $u(x) = -e^{-\alpha x}$, with a strictly positive α. An investor characterized by this exponential utility has to allocate an initial wealth W_0 between a risk-free and a risky asset. We assume a binomial uncertainty model, so that the risky asset has two possible gains (not returns) R_u and R_d, with probabilities π_u and π_d, respectively. Let q be the wealth allocated to the risky asset; it is possible to borrow cash as well to short-sell the risky asset. How does q change as a function of initial wealth W_0? Do you think that your utility function is exponential?

7.2 An investor endowed with an initial wealth $W_0 = 1$ (e.g., in euro) maximizes the expected value of the quadratic utility function $u(x) = ax - bx^2/2$, where x is the terminal wealth obtained by investing in n risky assets. Accordingly, the investor chooses a portfolio. Another investor, with a different initial wealth of $W_0 = K$, by optimizing the *same* utility function, chooses a different portfolio (in the sense that the asset weights are different).

- How do you explain the difference?

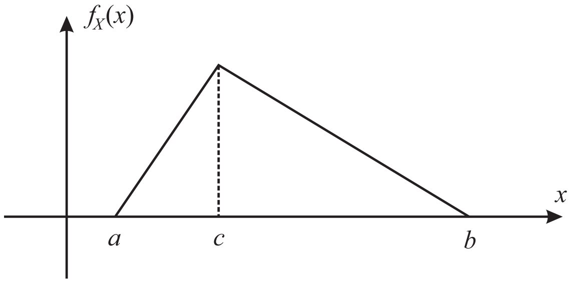

FIGURE 7.5 **PDF of a triangular distribution.**

- If the second investor changes the coefficient b to b', he finds the same portfolio as the first investor. What is the relationship between b and b'?

Note: This problem is borrowed from [13].

7.3 The value of your real estate property is \$100,000. In case of a fire, your property may be lost or damaged, depending on how severe the accident is. Let us assume that the following scenarios give the residual value of your property in the future, depending on the possible occurrence of a fire:

State	Residual value	Probability
ω_1	\$100,000	0.95
ω_2	\$50,000	0.04
ω_3	\$1	0.01

State ω_1 means that no accident occurred. Assume that your preferences are represented by a logarithmic utility depending on wealth, which is why we do not consider a residual value of \$0, but \$1. What is the maximum price that you would be willing to pay for an insurance guaranteeing coverage of any loss? Note: In the three states, the insurance will pay \$0, \$50,000, and \$99,999, respectively, so that the value of your property is preserved.

7.4 Consider random variables L_1 and L_2, modeling loss from two portfolios, and assume that they are jointly normal. Show that, in this case, value-at-risk is subadditive.

7.5 Figure 7.5 shows the probability density function (PDF) of a generic triangular distribution with support (a, b) and mode c. For such a distribution, expected value and variance are given by the following formulas:

$$\mathbb{E}[X] = \frac{a+b+c}{3},$$
$$\mathrm{Var}(x) = \frac{a^2+b^2+c^2-ab-ac-bc}{18}.$$

Say that the profit from a financial portfolio, with a holding period of a few weeks, has a triangular distribution with parameters (in €) $a = -75{,}000$, $b =$

55,000, and $c = 40{,}000$, so that the maximum possible loss is €75,000. Find V@R at level 95%. Note: The drawing of Fig. 7.5 is not in scale and is just meant as a qualitative hint.

7.6 Consider the following payoff distributions for two independent investment opportunities:[26]

Investment A		Investment B	
Payoff	Probability	Payoff	Probability
4	0.25	1	0.33
5	0.50	6	0.33
12	0.25	8	0.33

Compare the two alternatives in terms of stochastic dominance. *Hint:* You may plot the CDF, which is piecewise constant, and its integral, which is piecewise linear, for the two alternatives.

Further reading

- Decision-making under uncertainty is a topic of general interest, which is treated in different ways by different academic and practitioner communities. A thorough treatment of utility theory can be found, e.g., in [12], which is a treatment with a more economic flavor.
- There is an array of excellent books offering a treatment in a more financial vein, dealing with both utility theory and stochastic dominance. A concise, yet quite broad coverage of portfolio theory is offered in [11]. You may also see [7]. A more extensive treatment is offered in [9] or [10].
- The concept of coherent risk measure was introduced in [1].
- Risk measures are dealt with extensively in books with a more computational twist, especially stochastic programming. You may see [15], as well as [14]. A quite readable chapter on risk measures can also be found in [6].
- We have defined the concept of stochastic dominance, but portfolio optimization using this framework is more challenging than using utility functions or risk measures. See, e.g., [8] or [17].

Bibliography

1 P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[26] This is a numerical example borrowed from [7].

2 P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku. Coherent multiperiod risk adjusted values and Bellman's principle. *Annals of Operations Research*, 152:5–22, 2007.

3 Z. Bodie, A. Kane, and A. Marcus. *Investments* (9th ed.). McGraw-Hill, New York, 2010.

4 P. Brandimarte. *Quantitative Methods: An Introduction for Business Management*. Wiley, Hoboken, NJ, 2011.

5 P. Brandimarte. *Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics*. Wiley, Hoboken, NJ, 2014.

6 M. Capiński and E. Kopp. *Portfolio Theory and Risk Management*. Cambridge University Press, Cambridge, 2014.

7 J.-P. Danthine and J.B. Donaldson. *Intermediate Financial Theory* (3rd ed.). Academic Press, San Diego, CA, 2015.

8 D. Dentcheva and A. Ruszczyński. Portfolio optimization with stochastic dominance constraints. *Journal of Banking & Finance*, 30:433–451, 2006.

9 C. Gollier. *The Economics of Risk and Time*. MIT Press, Cambridge, MA, 2001.

10 J.E. Ingersoll, Jr. *Theory of Financial Decision Making*. Rowman & Littlefield, Totowa, NJ, 1987.

11 M.S. Joshi and J.M. Paterson. *Introduction to Mathematical Portfolio Theory*. Cambridge University Press, Cambridge, 2013.

12 D.M. Kreps. *Notes on the Theory of Choice*. Westview Press, Boulder, CO, 1988.

13 D.G. Luenberger. *Investment Science* (2nd ed.). Oxford University Press, New York, 2014.

14 G.C. Pflug and A. Pichler. *Multistage Stochastic Optimization*. Springer, Heidelberg, 2014.

15 G.C. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk*. World Scientific, London, 2007.

16 S.L. Savage. *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty*. Wiley, Hoboken, NJ, 2012.

17 A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, PA, 2009.