Chapter 3

Modeling the Objective Function

> If you do not know where you are going, every road will get you nowhere.
> – *Henry Kissinger*

> A goal without a plan is just a wish.
> – *Larry Elder*

The objective function of a mathematical program is what an optimization procedure uses to select better solutions over poorer solutions. For example, if the objective is to maximize profit, then the procedure tries to move in the direction of solutions that increase profit while still remaining feasible. But when the profit depends on a parameter that is uncertain (like prices tomorrow), then the notion of maximizing profit is no longer very simple.

3.1 Distribution of Outcomes

The broadest perspective you could take on this question is that your decision, once taken today, results in a distribution of outcomes. Your choice amounts to a choice of one distribution from a whole family of outcome distributions "parameterized" by your decision. If you ignore the randomness in the problem (as many do, but of course you are not one of them!), then your procedures will still select one distribution from this family—but you will be unable to control that choice. You need some way to inform the optimization process to select distributions with favorable characteristics.

> Making a decision can be viewed as choosing one particular outcome distribution over all others.

But what is a favorable characteristic of an outcome distribution? This is a question with no simple answer. There have been many scientific avenues of

inquiry into this issue and no one way of looking at the question has emerged, and many strange paradoxes remain. We will be content to indicate some practical concepts that can be used to select better outcome distributions over poor outcome distributions.

3.2 The Knapsack Problem, Continued

Let us first review the knapsack problem in its soft-constraint formulation:

$$\max_{x_i \in \{0,1\}} \quad \sum_{i=1}^{n} c_i x_i - d \sum_{s \in \mathcal{S}} p^s \left[\sum_{i=1}^{n} w_i^s x_i - b \right]_+ . \tag{3.1}$$

Consider a solution $\hat{x}$. Unless we have a large capacity, there will be a collection of sample points $s \in \mathcal{W}(\hat{x})$ representing scenarios where the combined weight of the selected items turns out to be larger than the maximum weight allowed. So the distribution of objective function values will be a random variable:

$$V_s(\hat{x}) = \begin{cases} \sum_i (c_i - dw_i^s) \hat{x}_i + db & \text{if } s \in \mathcal{W}(\hat{x}), \\ \sum_i c_i \hat{x}_i & \text{otherwise.} \end{cases}$$

To analyze a solution to a stochastic program, you will need to examine the distribution of the objective value and ask yourself: what are the features that we are concerned about?

Perhaps the most important practical features of the distribution are the expected value and the upper and lower quantiles, for example, the 10% and 90% quantiles. Are these what you expected? Should the penalty d or the weight limit b be adjusted?

The main point we wish to make here is that it is a big challenge to design an objective function that fully captures all the desired features of the outcomes, and it is very rare to get it right the first time!

Soft constraints partition the underlying sample space into favorable and unfavorable outcomes. It is perhaps worth examining this partition to learn whether this is what was intended.

Consider a knapsack problem with two different customers. If the total weight of accepted items from the two customers were very different, this might be a bad outcome. For example, let us say that items in the set O_1 come from the first customer class and items in the set O_2 come from the second customer class. We can track this difference by calculating the expected difference in the weights of the excluded items:

$$\sum_{s \in \mathcal{W}(\hat{x})} p^s \left| \sum_{i \in O_1} w_i^s \hat{x}_i - \sum_{i \in O_2} w_i^s \hat{x}_i \right| .$$

When this difference is too large, it might be bad for customer relations! Now, think for a minute—how could you change the problem to address this issue?

Consider also the expected contribution of each item to the objective function, which can be modeled as the profit for including the item less its expected contribution to an overweight situation:

$$\left(c_i - d \sum_{s \in \mathcal{W}(\hat{x})} p^s w_i^s \right) \hat{x}_i.$$

Slicing and dicing the contributions by item attributes may lead to important insights into other features that need to be controlled. The main point is that you should look carefully at the outcome distribution and verify whether its properties were intended.

3.3 Using Expected Values

The most commonly used way of comparing two outcome distributions is to compare the expected values. By "using expected values" we do not mean that you are using a single point (the mean) as the realization of the random parameters. Rather we mean that you are optimizing the expected value of the outcome distribution.

Sometimes we choose an expected value criterion because this is the simplest and most convenient approach. But in many cases it is also the right thing to do. Here we will outline the most common arguments for maximizing expected profit or minimizing expected cost.

3.3.1 You Observe the Expected Value

Where you face a situation that will be repeated over and over again, then simple repetition favors the expected value criterion.

For example, the news vendor of Sect. 1.2 makes a fixed order that will be used daily, say, for the next year. In such a case, even if there are severe variations in daily costs, it is still reasonable to minimize *expected* cost. The law of large numbers takes over. Your annual result may be so close to the expected cost that you would not care about the difference.

> If you are planning for outcomes that will be repeated many times, then it is likely correct to use expected value as the criterion for choosing an outcome distribution.

Before rejecting the mean value as an objective criterion, you should dig into the operational details of how the decisions will be used in operations. For example, even in the case where the news vendor revises the order on a weekly basis, the average of the weekly costs over a year will also approximate the mean value—even though the variation in the weekly costs may be quite large. (How would you go about verifying this statement?)

3.3.2 The Company Has Shareholders

Making decisions under uncertainty in a corporate setting has some special features that are the subject of a deep and extensive literature. We will base our discussion here on a paper by Froot and Stein [15]. The basic idea is that a public company has shareholders who themselves are making decisions under uncertainty about how many shares they wish to hold in which company. Very often shareholders want to be exposed to risk since risk also means opportunity. The question is what sort of risk management should be pursued by the company?

Let us suppose that the company faces a major internal decision with risky outcomes. For simplicity of argument, let us also assume that even in the worst of circumstances, the company will not go broke. (The issue of bankruptcy is discussed below.)

Here are a couple of questions:

- Would you be risk averse in this situation or go for the expected value?
- Would you be willing to buy insurance to avoid the worst outcomes?

The way to answer this argument, according to [15], is to put yourself in the shoes of one of your shareholders. Let us represent the shareholder by a wise lady who understands that most of the risk in your profit is caused by your technology choices. She understands that companies may have different solutions to the problem at hand, and your success (or failure) depends on which solution your customers end up preferring.

To hedge this uncertainty concerning technologies, she buys equally many shares of your competitor's company. From her perspective, there is no longer any risk—except those that relate to her exposure to the market for your and your competitor's products.

What will happen now if both companies recognize the risks they are facing? For example, suppose they both decided to insure their technology risks? Well, our wise lady still faces no market risk, but now, whichever technology wins, she gets less! She will not be very happy with this decision.

If only one of the companies reduces risks, then it is even worse. Possibly without knowing it, she now faces risks since the symmetry between the two companies is gone *and* the expected value of her investments has gone down. So to assure investors like her that she can safely invest in your company, she must be assured that you will not behave in a way that increases the risks for her.

A company with shareholders should maximize expected profit and
leave risk adjustments to the owners, except for those risks that can
be hedged more cheaply by the company itself.

Some risk-reducing measures are only available to companies, as a result of
taxation rules, perhaps, or as a result of access to markets. If such measures are
available to a company, then it can and should be risk averse in those respects.
It is crucial that investors realize that this is being done. Many companies
make statements to this respect in their public reporting. If you run into this
problem in a practical setting, it is wise to consult with financially trained
people. Our point is that you become aware that even in decisions where you
do not observe repeated outcomes, it may not be appropriate to reduce all
the risk you face.

Our wise lady's strategy is related to a very popular statistical arbitrage
called "pairs trading," in which investors take a short position in one com-
petitor and an offsetting long position in another. The investor is not exposed
to the success or failure of the underlying market but will make money on a
temporary deviation that supposedly should revert to the statistical long-run
mean. This kind of gamble is really a byproduct of how large corporations
are managed. There really is no way to reward management for outcomes
based on whether a given technology is good or bad. The actual practice is
to reward *steady earnings growth* of the sort that can only be achieved by
holding a diversified portfolio generating the products and services sold by
their sales forces. If two large companies hold diversified portfolios and com-
parable brands, then taking bets on statistical properties of their earnings is
a plausible arbitrage strategy.

3.3.3 The Project Is Small Relative to the Total Wealth of a Company or Person

Even if the variance of the income from an investment is very high relative
to the mean, this variability is usually not cause for concern if the numbers
happen to be small. When someone tells you: "The cost will be two or three
dollars, I am not sure which," you probably say: "OK, I don't mind." You
are not concerned despite the fact that one estimate is 50% higher than the
other.

On the other hand, if the same person tells you, "Oh, it will cost two or
three million dollars," you probably would hesitate, even if the ratio between
the two cost estimates is the same. Why? Probably because in the latter case,
the amounts are substantial relative to your total wealth.

Use expected value if the relative importance of the project is small.

This is actually a variant of the first argument about observing the expected value. If a project is small relative to a company's total wealth, then the company probably has a very large number of these projects, and it will observe the expected value (but now the expectation is over projects and not over outcomes within a single project).

If none of these arguments applies, then it is probably time for you to think properly about what risk actually means in your case and if you should be worried about it.

3.4 Penalties, Targets, Shortfall, Options, and Recourse

When soft constraints are incorporated into an objective function, the part of the objective function that models the soft constraints will have a certain shape. Again, look at the knapsack problem (3.1). The soft constraint part of the objective is the expected value of a piecewise linear function of the excess weight. It is zero below the capacity b and has slope d above the capacity. This piecewise linear function has many names and appears in many contexts.

3.4.1 Penalty Functions

In the language of linear programming, such functions are sometimes called *penalty functions*. The penalty function has two parts. One is the *target interval*, which in the knapsack case is the interval below the capacity, namely $(-\infty, b]$. The second part is the *penalty rate*, namely the rate at which the penalty accumulates as the target is missed. This rate may be an actual cost of responding to the penalty, but more usually it is a modeling approach that is used by the modeler to shape the outcome distribution. An example for the knapsack problem can be found in Fig. 3.1.
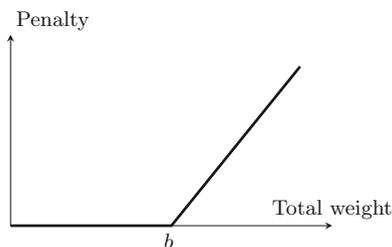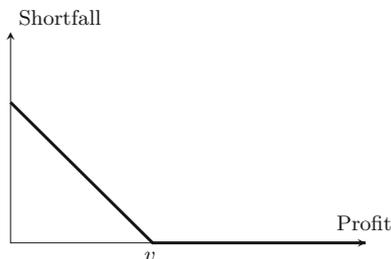

Fig. 3.1: Penalty function for knapsack problem

3.4.2 Targets and Shortfall

One can use penalty functions to indicate a desire to reach a target. Although there is a similarity to a penalty formulation of a soft constraint, a

Fig. 3.2: Example of shortfall function with a target v

target formulation is not really a soft constraint since the act of selecting a target reshapes the outcome distribution. If you prefer outcomes x^s above a given target v to outcomes below the target, then the simplest criterion that measures this preference is called the *shortfall measure*:

$$\sum_{s \in \mathcal{S}} p^s (v - x^s)^+. \tag{3.2}$$

Consider the function

$$h(y) = \begin{cases} 0 & \text{if } y < 0, \\ y & \text{if } y \geq 0. \end{cases} \tag{3.3}$$

This is a piecewise linear function with slope 0 below zero and slope 1 above zero. An illustration of a penalty function can be found in Fig. 3.2. It is not hard to see that

$$\sum_{s \in \mathcal{S}} p^s h(v - x^s) = \sum_{s \in \mathcal{S}} p^s (v - x^s)^+. \tag{3.4}$$

This function compares two outcome distributions by looking at the expected values over the region below the target v. So in the situation where x^s represents the return of an investment portfolio over some time horizon, then perhaps we would like to avoid outcomes with higher values of $h(v - x^s)$. If we are in a maximizing frame of mind (most investors are), then we could maximize the following expression:

$$\max_{x \in X} \left\{ \sum_{s \in \mathcal{S}} p^s x^s - \sum_{s \in \mathcal{S}} p^s h(v - x^s) \right\}, \tag{3.5}$$

where X is some set that constrains our portfolio choices. What are we maximizing here? We are maximizing a piecewise linear function, which, computationally, is not too hard. What would it do? Well, if we had two outcome distributions with similar means, this method of choosing outcomes would prefer the one with a larger *conditional* expectation over the outcomes that lie above the target v.
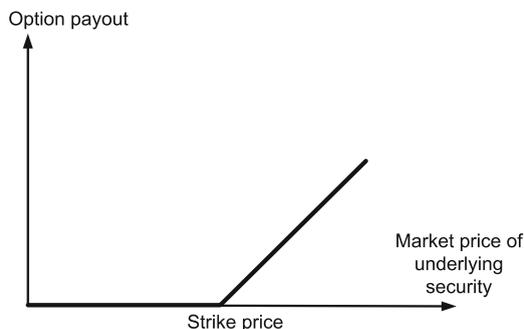
Fig. 3.3: Value of call option as a function of the price of the underlying
            security

These types of shortfall measures are very useful in applications involving
uncertainty. Target shortfall measures are a natural way to describe differences
in outcome distributions, and the optimization technology required to solve
them is readily available. We will find many examples of these in this book.

3.4.3 Options

Penalty functions model a cost that is incurred when certain underlying events
occur. In finance, a contract with terms that incur a cost or produce a payment
depending on the occurrence of a future event is called an *option.* A typical
type of option is a call option, which grants the owner the right to buy an
underlying security at a fixed price, called the strike price, at some fixed date
in the future.

The owner of the call option has a choice on the exercise date. If the
price of the underlying is above the strike price, then the owner can buy the
underlying security for the strike price and then sell it at the market price.
The owner's profit equals the difference between the market price and the
strike price. On the other hand, if the market price is below the strike, then
the owner need not do anything. The option payout is a function that is zero
below the strike and increases linearly with slope 1 above the strike, as in
Fig. 3.3.

Can you see that an option payout (Fig. 3.3) looks like a penalty function
(Fig. 3.1) with a target equal to the interval above the strike price and a rate
equal to one? Also note the similarity with the shortfall function (Fig. 3.2).
Many penalty formulations can be framed in terms of call and put options
because penalties are essentially invoked when a stochastic value goes above
or below a target.

3.4.4 Recourse

Another term that applies to penalty formulations is *recourse.* A recourse
model describes actions that lead to a future cost or benefit in response to
future events. In the case of a penalty formulation, the recourse model just

calculates the penalty (such models are called "simple recourse" in the literature). But a recourse action could be more complex. A recourse model can minimize the impact of a bad event using multiple technologies that are available to the decision maker but that may not be available to investors.

Investors use options to implement strategies for portfolio management. In the same way, decision makers may invest in recourse capabilities to improve their capabilities for managing uncertainty. In a sense, a stochastic program with recourse can be viewed as an option portfolio selection model. However, recourse is a concept that goes far beyond options. Recourse is modeled from the collection of possible actions and resources available to the decision maker, so in a sense, the use of recourse models allows decision makers to design their own options.

3.4.5 Multiple Outcomes

Typically in stochastic programming, you are concerned with multiple outcomes. In the foregoing example, we were concerned with maximizing the mean return and at the same time wanted to minimize the shortfall measure. The natural thing to do was to parameterize the problem:

$$\max_{x \in X} \left\{ \sum_{s \in \mathcal{S}} p^s x^s - \lambda \sum_{s \in \mathcal{S}} p^s h(v - x^s). \right\} \tag{3.6}$$

Varying the parameter gives an "efficient frontier" of solutions; $\lambda = 0$ gives the solution that maximizes expected return, and as λ grows higher and higher, it will give the solution that minimizes the shortfall measure.

This "multiobjective" style of optimization procedure is very common in stochastic programming. It allows us to describe multiple targets and objectives, and the optimization process generates efficient sets of solutions, each with different properties relative to the targets.

3.5 Expected Utility

The problem of choosing outcome probability distributions has a very deep technical literature that centers around the economic utility theory developed by von Neumann and Morgenstern [54]. The basic idea is that preferences between outcome distributions (following certain rules) can be modeled by choosing an outcome distribution that maximizes *expected utility*, where utility is modeled by a concave function of the outcomes. We will give just a brief outline here of the portfolio selection problem in finance since it is in finance that the basic assumptions of expected utility are likely to be satisfied.

Let us place ourselves in the realistic world of choosing to invest in corporations that are in effect managing portfolios of businesses. As observed

previously, one can anticipate some statistical regularity of outcomes that will be observed as dividend payments or changes in the market prices of company stock. Purchasing a single share of stock in company i will produce an annual return of $r_i(s)$, where s is a scenario parameter indicating the strength of the market returns modified by the idiosyncratic performance of management. (Just to be clear on the meaning of *return*, we will adopt the convention that a return less than 1.0 represents a loss and one greater than 1.0 a gain.)

Investing in a portfolio $x = (x_i, i \in I)$ of companies will therefore produce an outcome distribution of

$$s \mapsto \sum_{i \in I} x_i r_i(s), \text{ with probability } p(s). \tag{3.7}$$

Since companies are run by managers who are rewarded for earnings growth, it is quite likely that there is some statistical regularity to the dividend payments. Under the assumptions of *expected utility*, then, there exists a utility function $F(\cdot)$ such that the optimal choice of outcome distribution is given by maximizing expected utility:

$$\max_x \sum_{s \in \mathcal{S}} p^s F\left(\sum_{i \in I} x_i r_i(s)\right). \tag{3.8}$$

Now we ask the question—what should the utility function be? In the case of expected value optimization, the utility function is just the identity $F(R) = R$. Should we use the expected value criterion to choose an optimal collection of stocks? What are the other choices?

If our perspective is very long term (for the rest of our long lives, for example) and our objective is to simply take the money every year and *spend it*, then the expected value discussion applies: the variability over many many years will oscillate around the mean.

On the other hand, if we take the money and *reinvest it*, then the story is different. When the returns are identically distributed, the strategy that achieves the maximum wealth is the one that *maximizes the logarithm* of the return. (Of course, this is simply the mean of the exponential growth one achieves through reinvestment—so the mean wins out here, too!). This result is originally due to Kelly and has been developed in the stochastic programming context by Ziemba and his colleagues [41].

Of course, these objectives assume we know the distributions rather precisely. In fact, the distributional characteristics of investment returns change over time. The next section investigates an important tool used by portfolio managers to model the risks of investment.

3.5.1 Markowitz Mean-Variance Efficient Frontier

You have likely heard of the Markowitz criterion, in which the objective minimizes the variance for a given level of expected return [44]. Why is this so popular? Well, for one thing it uses observable statistics—the mean and

variance of financial returns are easily observable. But is it sensible? After all, variance penalizes both downside risk and upside risk. Is it reasonable for an investor to choose a criterion that minimizes the risk of going higher?

To answer this question, let us consider an investor who knows her statistics. For instance, she knows the mean return vector m and the variance-covariance matrix V. She also knows her von Neumann–Morgenstern and wants to choose her portfolio according to maximum utility. But what utility function? She goes to a Web site that offers to discover her utility by asking questions about one gamble after another. But she is really not sure about this at all. So many comparisons!

Along comes a slick stochastic programmer who offers to give her an *entire collection of optimal portfolios* to choose from. Every one will be optimal for some utility function. And up to a second-order approximation, every utility function will be represented. How does he do it?

He argues like this. Suppose your utility function was $F()$ and we knew how to find its optimal portfolio, namely $\hat{x}$ maximizes (3.8). Then of course we know its expected return, namely $\hat{R} = \sum_i m_i \hat{x}_i$. Expand the utility function to second order around this expected return:

$$F(R) \sim F(\hat{R}) + F'(\hat{R})(R - \hat{R}) + 1/2 F''(\hat{R})(R - \hat{R})^2. \tag{3.9}$$

Now find the maximum utility using the right side of the approximation instead of the left:

$$\max \sum_s p^s \left[F(\hat{R}) + F'(\hat{R}) \left(\sum_i r_i(s) x_i - \hat{R} \right) \right.$$
$$\left. + 1/2 F''(\hat{R}) \left(\sum_i r_i(s) x_i - \hat{R} \right)^2 \right]. \tag{3.10}$$

Without loss of generality, let us also restrict the search to those choices that satisfy

$$\hat{R} = \sum_i m_i x_i. \tag{3.11}$$

This does not specify the choice of x (it does narrow it down considerably, but let us keep going). The main point to keep in mind in the argument is that the choice of F determines $\hat{R}$. Now with this narrowing down, let us look carefully at the approximate utility maximization. First, note that the term $F(\hat{R})$ is fixed, so it will be ignored in the approximate maximization. Second, note that the second term disappears! This is because we are restricting our choices of x to those that lie on the mean hyperplane (3.11). Finally, note that the last term consists of the second derivative of a concave function (which is negative) multiplied by the variance of the return. When we clear the negative term from the objective, the maximization turns into a minimization and we are left with the problem of minimizing the variance subject to a constraint on the mean return.

It follows that the approximation is none other than a version of the mean-variance problem central to the Markowitz method:

$$
\begin{aligned}
&\min \quad \sum_{i,j} x_i V_{ij} x_j \\
&\text{such that } \hat{R} = \sum_i m_i x_i.
\end{aligned}
\tag{3.12}
$$

As we vary our choices of utility $F()$ we will also vary our choices of return $\hat{R}$. It follows that all our choices will lie on the *efficient frontier* of solutions that minimizes variance for a given level of mean return. This is the approach originally formulated by Markowitz [44]. The mean-variance efficient frontier does in fact present our investor with a collection of points that a utility-maximizing investor would choose, up to a second-order approximation.

How good is the approximation? Well, this is something you can try for yourself. Find some tables of annual returns of large corporations over the past 20 years, calculate the means and variances, and answer for yourself: how good is the second-order approximation to your favorite utility function, say, the logarithm? You will find that it is pretty close. After developing this argument (in [34]) we asked this question. For the logarithm function it seemed like the second-order approximation was very sensible for absolute returns in the range of 75–400%—which is one very good reason for the popularity of the Markowitz method over the 60 years since its discovery.

The other reasons are that the parameters are quite easily observed in the marketplace. The covariance matrix and mean can be constructed by observing a time series of annual returns. The sequence of observations is viewed as samples drawn from the distribution of future returns. Standard statistical calculations can be used to provide appropriate estimates. It appears that there are long-term cycles in *market* volatility; however, the same cannot be said about variances and correlation terms for individual stocks—partly because there is less data to estimate them and partly because the relative performance of company stock prices depends on so many factors. However, the actual performance of the mean-variance model over time is much more sensitive to the estimation error in the fundamental parameters, most especially the mean.

At this point we will close this discussion. This is not a book about statistical arbitrage in financial markets; it is a book about modeling choices under uncertainty. We hope we have conveyed to you some of the flavor and language of expected utility and the mean-variance approach as it is applied in investing.

The interested reader can go much further, of course. However, in the end, we would like you to be aware that many market practioners, on the basis of much experience, do not believe that past data inform the future behavior of prices. Rather, the basic reason for prices to move one way or another is due to supply and demand—and in the securities market the dynamics of supply
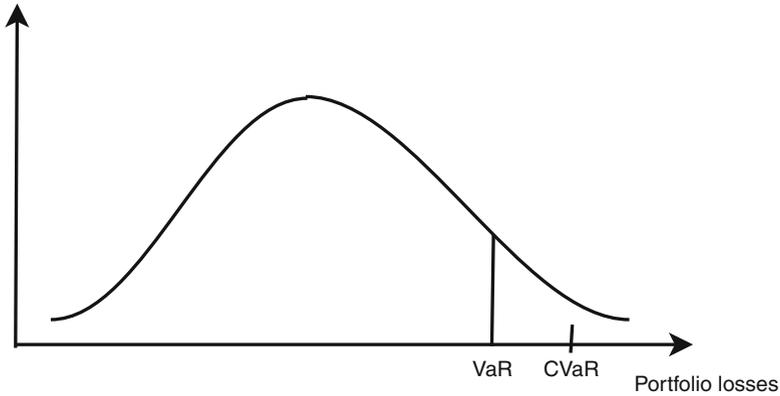
Fig. 3.4: Density function, VaR, and CVaR for potential losses. VaR corresponds to a chosen quantile, while CVaR is the expected loss above VaR

and demand are affected at times by overwhelming optimism and at times by overwhelming pessimism and always amplified by leverage. This brings us to our next topic.

3.6 Extreme Events

The 50-50-90 rule: Any time you have a 50-50 chance of getting something right, there's a 90% probability you'll get it wrong.
– *Andy Rooney*

Sometimes extremely bad things happen. Asteroids strike, virulent diseases break out, markets crash, products fail. Models with uncertainty must consider the consequences of extreme events.

In the financial industry, for example, regulators require some institutions to estimate the upper tail of the loss distribution and to hold reserves proportional to these loss estimates. The intention of the regulators is to force the industry to model the worst-case extreme losses of their portfolios and to penalize extremely risky positions by forcing them to hold safe, but low-yielding, securities in proportion to the extreme-loss potential. The upper quantile of potential losses is called *Value at Risk* or VaR. A related, increasingly popular, and, in our view, preferable statistic is called the *Conditional Value at Risk* or CVaR—which is the expected value of the losses that fall above the VaR. This is illustrated in Fig. 3.4.

The first step in analyzing extreme risks is to decide what to analyze. In the preceding investment problem, we could decided to set a *benchmark* return. For example, we could have chosen a level R_s of returns (keep in mind that a

return below 1.0 means that our investment has lost value) that represents an absolute threshold below which we do not want to go. Then our *losses* relative to the benchmark are

$$L^s(x) = R_s - \sum_i x_i r_i^s \quad \text{for } s \in \mathcal{S} \text{ with probability } p_s. \tag{3.13}$$

The VaR for the losses L corresponding to a quantile Q is calculated by sorting the losses from lowest to highest and counting from the lowest loss up until you have counted a proportion $Q\%$ of all scenarios. Here is a formula that says the same thing but is switched around to highlight something that will become important a bit later on:

$$\text{VaR}(x; Q) = \inf_L \text{ such that } \#\left\{ s \in \mathcal{S} \text{ with } L^s(x) - L \geq 0 \right\} \geq \frac{1-Q}{100}, \tag{3.14}$$

where the number sign $\#$ in front of the set indicates the number of points in the set. The risk-level $\text{VaR}(x; Q)$ divides the sample space of losses into two parts: good and bad. The good points are the $Q\%$ that have losses below the $\text{VaR}(x; Q)$ and the bad points are the $(100\text{–}Q)\%$ that have losses greater than the $\text{VaR}(x; Q)$.

In banking and finance one says "our 99% VaR is USD 0.5B," and it is understood by bank regulators that the probability that the bank's position will suffer a loss greater than $500 million is only 1%. These statements are taken quite seriously. In fact, the bank must demonstrate to the regulators, by reconstructing their positions backward in time, that their past 200 daily 99% VaR loss estimates were violated no more than twice.

But can you see that VaR behaves a bit strangely as a function of x? Try this thought experiment: take a position x_i in the security with the absolute worst losses and make a small increase in it. A good risk measure should increase when you do more bad things. Is the VaR guaranteed to change? No, it is not, and this is why it is a controversial measure.

The CVaR is defined as the expected value over the quantile in definition (3.14) of VaR. Given that expected values are linear, one would anticipate that a risk measure based on CVaR will behave as a risk measure should. However, there is the complicating matter of locating the support of this expected value, which behaves badly, as we have seen. A relatively new result by Rockafellar and Uryasev shows that CVaR is indeed a well-behaved risk measure. They proved that its value is given by the following stochastic program:

$$\text{CVaR}(x; Q) = \inf_L \left\{ L + \frac{1}{1-Q} \sum_s p^s \max[0, L^s(x) - L] \right\}. \tag{3.15}$$

Do you see that this calculates expected value over the same set that we counted points over for the calculation of VaR?

In usage, VaR and CVaR have different purposes. Both are measures of the tail behavior of an outcome distribution. VaR tells you something about where the tail is supported, whereas CVaR gives you more information about how the distribution behaves over the tail.

Regulators prefer VaR because they have more confidence in the statement "only 5% of losses are greater than VaR" than they do in the statement "the average of the worst 5% of losses is CVaR." The second statement requires a model of the tail distribution, which by necessity must always be based on a sample of very few observations. People can disagree about tail distributions, but the definition of the location of the tail is usually pretty well estimated. Regulators will prefer to stick with estimates for which there is broad agreement.

On the other hand, decision makers should prefer CVaR, for two reasons. First, precisely because it does require a model of the tail distribution, the inclusion of CVaR will force decision makers to think about how bad things can get.

We can think of CVaR as being just like a target shortfall measurement, but where the target is specified in terms of the quantile instead of some arbitrary target. Of course we could just specify the target to be high enough to be "in the tail." The advantage of using CVaR is just the fact that we do not have to guess beforehand where these tail values are.

Sometimes it is a good idea to use a different distribution for extreme events. For example, on trading floors, the traders do not use the same distributions as the risk managers. Risk managers have different objectives and different time horizons. A trader about to make a quick killing should not be concerned about surviving a low-probability market crash over the next 30 days. It is the risk manager's job to worry about that. The risk manager does not do this by influencing the trader's models. (It is hard to imagine a trader accepting any kind of modeling advice from a risk manager!) Risk managers do their job by limiting trader access to the firm's capital. It is entirely reasonable that the risk manager and the trader use different distributions to achieve their objectives.

Extreme event modeling in finance is gradually being extended by regulators to cover so-called operational risks. This covers things like power failures at data centers, losses due to failures of internal controls, rogue traders (and risk managers), and so forth. The financial system is so interlinked and the traded volumes so great that an operation breakdown, say, due to a power failure at a major bank's data center, could have widespread financial and economic consequences for weeks afterward.

Companies in industrial sectors with large extended supply chains are also beginning to model the tail distributions of bad events, operational and otherwise. Examples of bad events could be the bankruptcy of a key supplier or creditor, quality control failures internally or of key suppliers, a major product liability lawsuit, improper release of private information of consumers, earthquake damage to a key electric power supplier, or losses in financial portfolios that back employee insurance programs or working capital.

3.7 Learning and Luck

> I'm a great believer in luck, and I find the harder I work, the more I
> have of it.
> – *Thomas Jefferson*

Learning is a subject of its own, and we will not have a general discussion
of the issue here. However, since such things as learning organizations and
organizational memory are preached by many consultants, we would like to
mention but one issue that is relevant for this book. Learning presumably
implies that from the outcome of a decision we wish to become wiser. Next
time we make a similar decision we wish to either make a better decision
(if this one was not so good) or an equally good one (if this one was good).
For this to make any sense, there must be a causal relationship between what
we did and what happened.

If you walked backward into the store where you bought the winning
lottery ticket, you might think (many certainly do) that this is good for win-
ning, and you might choose to do the same in the future. We all know that this
is nonsense, and we call it superstition, not learning. If, instead, you have a
model you use for making some decision, and the decision leads to, say, a very
good outcome, do you then know it was a good decision and a good model?
Is there a logical connection between the ex-post observation and what you
did? Let us again exaggerate a bit. Assume there are two gambles, both with
the same price for taking part and both with the same winning prize. Assume
in one gamble there is a 10% chance of winning the prize, in the other 90%.
Assume you play the game with a 10% chance and, voila, you win! Does that
mean you made a good decision? Of course not. Is there anything to learn
from the ex-post observation? No, there is not. You were simply stupid and
lucky. We can say that because there is an ex-ante evaluation that in this
simple case tells us that what you did was stupid. The fact that you were
lucky does not change that.

So what is learning? We will not answer that question. But be aware
that learning is a difficult issue in a random environment. In genuine decision
contexts, there is a causal but stochastic relationship between what you do
and the consequences. But learning can be hard because you must separate
luck from cleverness.

But do we care? Maybe for our own decisions we often do not. We prefer
to value our decisions based on ex-ante analysis, looking for decisions that
according to our own utility function maximizes expected utility. But there is
a context where we really are interested. If you are to engage a consultant, he
probably sells himself based on his track record (sounds great, does it not?).
Here we are interested in ex-post learning. Is he really good or just lucky?
Maybe he looks so good because he takes too many chances, and for that
reason he is definitely not the one we want. Some say that if your broker
makes a lot of money for you over a reasonable period of time, fire him! He
takes too big risks.

Springer