Chapter 6
Risk Measurement Principles

In this chapter, we take a close look at the principles of risk measurement. We argue that it is natural to quantify the riskiness of a position in monetary units so that the measurement of the risk of a position can be interpreted as the size of buffer capital that should be added to the position to provide a sufficient protection against undesirable outcomes. In the investment problems in Chap. 4, variance was used to quantify the riskiness of a portfolio. However, variance, being just the expected squared deviation from the mean value, does not differentiate between good positive deviations and bad negative deviations and cannot easily be translated into meaningful monetary values unless the future value we consider is close to normally distributed. The risk premium considered in Chap. 5 is more natural than the variance as a summary of the riskiness and potential reward of a position. However, the risk premium is difficult to use effectively to control the risk taking of a financial institution or to determine whether the aggregate position of a company or business unit is acceptable from a risk perspective. In this chapter, we will present measures of risk, including the widely used value-at-risk and expected shortfall, analyze their properties, and evaluate their performance in a large number of examples.

6.1 Risk Measurement

We now turn to the topic of how to measure risk. Consider two times, time 0, which is now, and a future time $\Delta t > 0$. We may choose to measure time in units of Δt and therefore take the future time to be 1.

Let V_1 represent the random value at time 1 of a portfolio. The precise meaning of portfolio is left unspecified but may include assets, liabilities, and any kind of contract that can be assigned a monetary value. To measure the risk of the portfolio, we analyze the probability distribution of V_1. The probability distributions assigned to V_1 are likely to vary among a group of individuals or organizations for which the

H. Hult et al., *Risk and Portfolio Analysis: Principles and Methods*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-1-4614-4103-8_6,

future portfolio value is of relevance. Moreover, the way the probability distribution of V_1 is transformed into a measurement of the riskiness of the portfolio may depend on the context.

An asset manager, whose main objective is to generate profits while controlling the risk of and size of losses, needs to consider the whole range of possible outcomes of V_1 together with possible externally imposed risk constraints and profitability requirements. A risk controller analyzes the part of the distribution of V_1 corresponding to unfavorable outcomes. In particular, the portfolio may be considered acceptable by the risk controller but not by the asset manager if it is not likely to produce a good return. Similarly, a portfolio that has good potential of producing high returns may be unacceptable to the risk controller who finds that the probabilities of large losses are too high and have been overlooked (or ignored) by the asset manager.

A regulator of a finance or insurance market wants to impose rules on risk taking that on the one hand prevents banks or insurance companies from taking too much risk, and thereby threatening financial stability, but on the other hand allows companies to be profitable. The rules must enable the supervisory authority to classify the overall position of a company as either acceptable or unacceptable. Moreover, the supervisor must be able to inform a company with an unacceptable position of suitable actions to obtain an acceptable position, for instance, the minimum additional capital that the company must raise and invest prudently in order to be allowed to continue its business.

Many properties of a portfolio can be understood in terms of the probability distribution (e.g., the density function or distribution function) of its future value V_1. However, probability distributions are difficult objects to compare. Therefore, it is tractable to come up with a good way to summarize, from a risk measurement perspective, the entire probability distribution in a single number. We now discuss how this can be done.

Suppose there is a reference instrument with percentage return R_0 from time 0 to 1. The precise meaning of the reference instrument may depend on the context in which we are quantifying risk. For simplicity, here we take it be risk-free zero-coupon bonds maturing at time 1. If B_0 is the current spot price of the bond with face value 1 at time 1, then $R_0 = 1/B_0$ is the percentage return on the risk-free zero-coupon bond.

Consider a linear vector space $\mathbb{X}$ of random variables X representing the values at time 1 of portfolios. We denote by ρ a function that assigns a real number (or $+\infty$) to each X in $\mathbb{X}$, representing a measurement of the risk of X. The number $\rho(X)$ is interpreted as the minimum capital that needs to be added to the portfolio at time 0 and invested in the reference instrument in order to make the position acceptable. If $\rho(X) \leq 0$, then X is the value at time 1 of an acceptable portfolio; no capital needs to be added. In principle, a risk measure ρ could assign different values to two equally distributed future portfolio values X_1 and X_2. Throughout the book, we will only consider risk measures ρ for which $\rho(X)$ depends on X only through its probability distribution.

Next we list and comment upon some properties that have been proposed as natural requirements for good risk measures.

Translation invariance. $\rho(X + cR_0) = \rho(X) - c$ for all real numbers c.

This property says that adding a certain amount c of cash (and buying zero-coupon bonds for this amount) will reduce risk by the same amount. In particular, for an unacceptable portfolio X, adding the amount $\rho(X)$ makes the position acceptable: $\rho(X + \rho(X)R_0) = \rho(X) - \rho(X) = 0$.

Monotonicity. If $X_2 \leq X_1$, then $\rho(X_1) \leq \rho(X_2)$.

This property says that if the first position has a greater value than the second position at time 1 for sure, then the first position must be considered less risky. A risk measure satisfying the properties translation invariance and monotonicity is called a monetary measure of risk.

It is often suggested that a risk measure should reward diversification. Loosely speaking, it is wise not to put all your eggs in the same basket. The following property describes how diversification should be rewarded.

Convexity. $\rho(\lambda X_1 + (1-\lambda)X_2) \leq \lambda\rho(X_1) + (1-\lambda)\rho(X_2)$ for all real numbers λ in $[0, 1]$.

In particular, if $\rho(X_1) \leq \rho(X_2)$ and ρ has the convexity property, then

$$\rho(\lambda X_1 + (1-\lambda)X_2) \leq \lambda\rho(X_1) + (1-\lambda)\rho(X_2) \leq \rho(X_2).$$

For example, investing a fraction of the initial capital in one stock and the remaining capital in another stock, rather than everything in the more risky stock, reduces the overall risk. A risk measure satisfying the properties translation invariance, monotonicity, and convexity is called a convex measure of risk.

Normalization. $\rho(0) = 0$.

The normalization property says that it is acceptable not to take any position at all. Note that convexity and normalization imply that for λ in $[0, 1]$

$$\rho(\lambda X) = \rho(\lambda X + (1-\lambda)0) \leq \lambda\rho(X),$$

which in turn implies that for $\lambda \geq 1$

$$\lambda\rho(X) = \lambda\rho\left(\frac{1}{\lambda}\lambda X\right) \leq \lambda\frac{1}{\lambda}\rho(\lambda X) = \rho(\lambda X).$$

We conclude that the risk increases at least linearly in the size of the position. A strict inequality for large λ would reflect the well-known difficulty of selling off a large position within a short amount of time without affecting the price too much.

Positive homogeneity. $\rho(\lambda X) = \lambda\rho(X)$ for all $\lambda \geq 0$.

This property means that if we double the size of the position, then we double the risk. Moreover, taking $\lambda = 0$ we find that $\rho(0) = 0$, i.e., the positive homogeneity property implies the normalization property.

Subadditivity. $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$.

This property says that diversification should be rewarded. A bank consisting of two units should be required to put aside less buffer capital than the sum of the buffer capital for the two units considered as separate entities. In particular, if the regulator enforces the use of a subadditive risk measure, then it does not encourage companies to break up into parts in order to reduce the buffer capital requirement. Note that convexity together with positive homogeneity implies subadditivity.

A risk measure ρ satisfying the properties of translation invariance, monotonicity, positive homogeneity, and subadditivity is called a coherent measure of risk. Whereas a coherent risk measure is also a convex risk measure, a convex risk measure need not be coherent.

It may seem unintuitive at first to define the risk measure ρ on the value of a portfolio at time 1. Suppose $R_0 = 1$ and consider a fund manager who at time 0 invests $V_0 = \$10$ million in a giveaway portfolio with a value of $V_1 = \$1$ million .at time 1 (\$9 million is given away). If the risk measure ρ satisfies the translation invariance and normalization properties, then the risk measure applied to the future value of the portfolio yields $\rho(\$1 \text{ million}) = -\$1$ million, which corresponds to an acceptable investment. How can giving away money be an acceptable investment? The explanation lies in the interpretation of the future value X. Let us consider two stylized cases. In the first case, the fund manager is managing his own money, $X = V_1$, and there is nothing unacceptable about letting the fund manager give away some or all of his capital. In the second case, the money of the fund belongs to the fund's investors. In this case, the initial capital should be viewed as a liability to the fund's investors and $X = V_1 - V_0$. Therefore, $\rho(X) = \$9$ million, which corresponds to an unacceptable investment.

Example 6.1 (Solvency capital requirement). In the Solvency II framework, which is a regulatory framework for the insurance industry, a company is considered solvent if $\rho(A_1 - L_1) \leq 0$, where A_1 and L_1 are the values of its assets and liabilities 1 year from now and ρ a monetary (translation invariant and monotone) risk measure. It is quite common to illustrate the solvency graphically in terms of a picture of the balance sheet of the insurance company with the current value of assets to the left and the current value of liabilities to the right, and with the insurer being solvent if the height of the left column exceeds that of the right column (Fig. 6.1).

Let A_0 be the current market value of the assets, and let L_0 be the current market value (or best estimate) of the liabilities. Since ρ is translation invariant, we may write

$$\begin{aligned}\rho(A_1 - L_1) &= \rho([A_0 - L_0]R_0 + [A_1 - A_0 R_0] - [L_1 - L_0 R_0]) \\ &= L_0 - A_0 + \rho(\underbrace{[A_1 - A_0 R_0] - [L_1 - L_0 R_0]}_{\Delta}).\end{aligned}$$

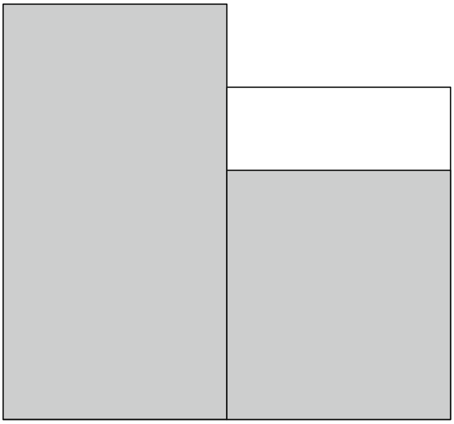

Fig. 6.1 Balance sheet. *Left*: present value of assets (*gray*). *Right*: present value of liabilities (*gray*). The solvency capital requirement is illustrated in *white*. The company is solvent if the present value of the assets is greater than the present value of the liabilities plus the solvency capital requirement

The quantity $\rho(\Delta)$ is called the solvency capital requirement and is denoted by SCR. A portfolio with a future value $A_1 - L_1$ is acceptable if $\rho(A_1 - L_1) \leq 0$, which is equivalent to $A_0 \geq L_0 + \text{SCR}$. The latter says that the current value of the assets exceeds the current value of the liabilities plus the solvency capital requirement. The balance sheet illustration of solvency may give the false impression, if not correctly interpreted, that solvency is about current asset and liability values, whereas solvency is really about future asset and liability values.

Example 6.2 (An absolute lower bound). Suppose that acceptable portfolios are those that are certain not to be below a fixed number c. This gives the risk measure

$$\rho(X) = \min\{m : mR_0 + X \geq c\}.$$

Define x_0 to be the smallest value that X can take (if no such value exists, then take x_0 to be the largest value smaller than all the values that X can take), and notice that

$$\rho(X) = \min\{m : mR_0 + X \geq c\} = \frac{c - x_0}{R_0}, \tag{6.1}$$

i.e., the discounted difference between the required capital c at time 1 and the worst possible outcome for the value of the portfolio at time 1. In particular, we note that if the portfolio contains short positions in some asset with an unbounded value at time 1 so that $x_0 = -\infty$, then $\rho(X) = +\infty$.

We claim that the risk measure ρ given by (6.1) is a convex measure of risk. To verify this claim, we need to show the translation invariance, monotonicity, and convexity. Translation invariance is shown by noticing that for all real numbers a,

$$\rho(X + aR_0) = \frac{c - (x_0 + aR_0)}{R_0} = \rho(X) - a.$$

To show monotonicity, we notice that if $X_2 \leq X_1$, then the corresponding lower bounds satisfy $x_{02} \leq x_{01}$, and therefore

$$\rho(X_1) = \frac{c - x_{01}}{R_0} \leq \frac{c - x_{02}}{R_0} = \rho(X_2).$$

Finally, we verify that the convexity property holds. If X_1 and X_2 have lower bounds x_{01} and x_{02}, then, for $\lambda \in [0, 1]$, the corresponding lower bound y_0 for $Y = \lambda X_1 + (1 - \lambda)X_2$ is greater than or equal to $\lambda x_{01} - (1 - \lambda)x_{02}$. Therefore,

$$\begin{aligned}\rho(\lambda X_1 + (1 - \lambda)X_2) &= \frac{c - y_0}{R_0} \\ &\leq \frac{c - \lambda x_{01} - (1 - \lambda)x_{02}}{R_0} \\ &= \lambda \frac{c - x_{01}}{R_0} + (1 - \lambda)\frac{c - x_{02}}{R_0} \\ &= \lambda\rho(X_1) + (1 - \lambda)\rho(X_2).\end{aligned}$$

Example 6.3 (Mean–variance risk measures). Consider portfolios whose future values X have finite variances and a risk measure of the form

$$\rho(X) = -\operatorname{E}[X/R_0] + c\sqrt{\operatorname{Var}(X/R_0)}, \quad c > 0. \tag{6.2}$$

By standard properties of the expected value and variance, it follows that ρ is translation invariant and positively homogeneous. Moreover, ρ is subadditive. This follows from the fact that

$$\begin{aligned}\operatorname{Var}(X_1 + X_2) &= \operatorname{Var}(X_1) + \operatorname{Var}(X_2) + 2\operatorname{Cor}(X_1, X_2)\sqrt{\operatorname{Var}(X_1)\operatorname{Var}(X_2)} \\ &\leq \left(\sqrt{\operatorname{Var}(X_1)} + \sqrt{\operatorname{Var}(X_2)}\right)^2,\end{aligned}$$

i.e., the standard deviation of the sum is less than or equal to the sum of the standard deviations for the two terms. Since ρ is both positively homogeneous and subadditive, it is also convex. However, the monotonicity condition is in general not satisfied, so ρ in (6.2) is not a convex measure of risk. The following example illustrates the lack of monotonicity. Let $X_1 = -R_0$ with probability one and let X_2 be a random variable that may take the values R_0 and $-R_0$, each with probability $1/2$. Then $X_1 \leq X_2$, and if $c > 1$ in (6.2), then

$$\rho(X_2) = c\sqrt{\operatorname{Var}(X_2/R_0)} = c > 1 = -\operatorname{E}[X_1/R_0] = \rho(X_1).$$

The lack of monotonicity is a serious flaw and limits the use of the mean–variance risk measure. However, for normally distributed random variables the mean–variance risk measure is canonical. If X is normally distributed, then we may write $X \stackrel{d}{=} \mathrm{E}[X] + \sqrt{\mathrm{Var}(X)}Z$, where Z is a standard normally distributed random variable. For any translation-invariant, positively homogeneous risk measure ρ we find that

$$\begin{aligned}\rho(X) &= \rho(\mathrm{E}[X] + \sqrt{\mathrm{Var}(X)}Z) \\ &= -\mathrm{E}[X/R_0] + \sqrt{\mathrm{Var}(X/R_0)}R_0\rho(Z).\end{aligned}$$

We conclude that as long as X is normally distributed, any translation-invariant, positively homogeneous risk measure satisfies the defining property (6.2) of mean–variance risk measures.

6.2 Value-at-Risk

The value-at-risk (VaR) at level $p \in (0, 1)$ of a portfolio with value X at time 1 is

$$\mathrm{VaR}_p(X) = \min\{m : \mathrm{P}(mR_0 + X < 0) \leq p\}, \tag{6.3}$$

where R_0 is the percentage return of a risk-free asset. In words, the VaR of a position with value X at time 1 is the smallest amount of money that if added to the position now and invested in the risk-free asset ensures that the probability of a strictly negative value at time 1 is not greater than p.

From (6.3) we see that $X \geq 0$ implies that $\mathrm{VaR}_p(X) \leq 0$. In order for VaR_p to be a sensible choice of risk measure for typical asset portfolios with mainly long positions, it is common to take the following view: at the current time 0 one starts from scratch and takes a risk-free loan of size V_0 (which is the current portfolio value), uses the capital to purchase the asset portfolio, and ends up with the net value $X = V_1 - V_0R_0$ at time 1. Therefore, the portfolio is classified as acceptable if the difference between the actual future portfolio value and the value that would be obtained by instead investing the current portfolio value in a risk-free asset is VaR_p-acceptable.

Before investigating the properties of VaR we first need to make sure that the minimum in (6.3) is attained so the definition really makes sense. To this end, note that

$$\begin{aligned}&\{m : \mathrm{P}(mR_0 + X < 0) \leq p\} \\ &\quad = \{m : \mathrm{P}(-X/R_0 > m) \leq p\} \\ &\quad = \{m : 1 - \mathrm{P}(-X/R_0 \leq m) \leq p\} \\ &\quad = \{m : \mathrm{P}(-X/R_0 \leq m) \geq 1 - p\}.\end{aligned} \tag{6.4}$$

Since a distribution function F is right continuous ($F(x) \downarrow F(x_0)$ as $x \downarrow x_0$) and increasing, $\{m : F(m) \geq 1 - p\} = [m_0, \infty)$ for some m_0, and therefore there exists a smallest element.

Set $L = -X/R_0$. If $X = V_1 - V_0 R_0$ is the net gain from the investment, where the current portfolio value V_0 is viewed as a liability, then $L = -X/R_0 = V_0 - V_1/R_0$ has a natural interpretation as the discounted loss. The identities in (6.4) give an alternative (equivalent) formulation of $\text{VaR}_p(X)$ in terms of L:

$$\text{VaR}_p(X) = \min\{m : \text{P}(L \leq m) \geq 1 - p\}. \tag{6.5}$$

We may interpret $\text{VaR}_p(X)$ as the smallest value m such that the probability of the discounted portfolio loss $L = -X/R_0$ being at most m is at least $1 - p$. Expressed differently, $\text{VaR}_p(X)$ is the smallest amount of money that, if put aside and invested in a risk-free asset at time 0, will be sufficient to cover a potential loss at time 1 with a probability of at least $1 - p$. Commonly encountered values for p are 5%, 1%, and 0.5%, which shows that $\text{VaR}_p(X)$ describes (to some extent) the right tail of the probability distribution of the discounted loss L. The length in physical time of the time period over which the discounted loss is modeled is often taken to reflect the time it may take to move out of an unfavorable position in the face of adverse price movements. In market risk measurement (e.g., stocks, bond and financial derivatives), the length of the time period is typically 1 day or 10 days, whereas 1 year is typical for credit and insurance risk measurement (e.g., retail or corporate loans or the aggregate value of assets and liabilities of an insurance company).

In statistical terms, $\text{VaR}_p(X)$ is the $(1 - p)$-quantile of L. The u-quantile of a random variable L with distribution function F_L is defined as

$$F_L^{-1}(u) = \min\{m : F_L(m) \geq u\},$$

and F_L^{-1} is just the ordinary inverse if F_L is strictly increasing. If F_L is both continuous and strictly increasing, then $F_L^{-1}(u)$ is the unique value m such that $F_L(m) = u$. For a general F_L, the quantile value $F_L^{-1}(u)$ is obtained by plotting the graph of F_L and setting $F_L^{-1}(u)$ to be the smallest value m for which $F_L(m) \geq u$. With this notation it follows that

$$\text{VaR}_p(X) = F_L^{-1}(1 - p). \tag{6.6}$$

To better understand the properties of the risk measure VaR_p, we first study the quantile function in more detail. We denote the uniform distribution on the interval $(0, 1)$ by $U(0, 1)$, i.e., the probability distribution of a random variable U satisfying $\text{P}(U \leq u) = u$ for u in $(0, 1)$.

Proposition 6.1. *Let F be a distribution function on $\mathbb{R}$. Then:*

(i) $u \leq F(x)$ *if and only if* $F^{-1}(u) \leq x$.
(ii) *If F is continuous, then* $F(F^{-1}(u)) = u$.
(iii) *(Quantile transform) If U is $U(0,1)$-distributed, then* $\text{P}(F^{-1}(U) \leq x) = F(x)$.
(iv) *(Probability transform) If X has distribution function F, then $F(X)$ is $U(0,1)$-distributed if and only if F is continuous.*

Proof. (i): Suppose $F^{-1}(u) \leq x$. By definition, $F(F^{-1}(u)) = F(\min\{y : F(y) \geq u\}) \geq u$. Since F is nondecreasing, $u \leq F(F^{-1}(u)) \leq F(x)$. Suppose now that $u \leq F(x)$. Since F is nondecreasing, $F^{-1}(F(x)) = \min\{y : F(y) \geq F(x)\} \leq x$. Since F^{-1} also is nondecreasing, $F^{-1}(u) \leq F^{-1}(F(x)) \leq x$.

(ii): As in (i) we have $u \leq F(F^{-1}(u))$. Take $y < F^{-1}(u)$ and note that by (i) this is equivalent to $F(y) < u$. Now, if $F(y) = \mathrm{P}(X \leq y) < u$ for all $y < F^{-1}(u)$, then $\mathrm{P}(X < F^{-1}(u)) \leq u$. Then

$$\begin{aligned} u \leq F(F^{-1}(u)) &= \mathrm{P}(X < F^{-1}(u)) + \mathrm{P}(X = F^{-1}(u)) \\ &\leq u + \mathrm{P}(X = F^{-1}(u)). \end{aligned}$$

The continuity of F implies that $\mathrm{P}(X = F^{-1}(u)) = 0$. We conclude that $u = F(F^{-1}(u))$.

(iii): $U \leq F(x)$ if and only if $F^{-1}(U) \leq x$ by (i). Hence, $\mathrm{P}(F^{-1}(U) \leq x) = \mathrm{P}(U \leq F(x)) = F(x)$.

(iv): Suppose F is continuous. By the quantile transform and (ii),

$$\mathrm{P}(F(X) = u) = \mathrm{P}(F(F^{-1}(U)) = u) = \mathrm{P}(U = u) = 0.$$

Hence, by (i)

$$\begin{aligned} \mathrm{P}(F(X) \leq u) &= \mathrm{P}(F(X) < u) + \mathrm{P}(F(X) = u) \\ &= \mathrm{P}(F(X) < u) \\ &= 1 - \mathrm{P}(F(X) \geq u) \\ &= 1 - \mathrm{P}(X \geq F^{-1}(u)) \\ &= \mathrm{P}(X < F^{-1}(u)) \\ &= F(F^{-1}(u)). \end{aligned}$$

It now follows from (ii) that $F(X)$ is $U(0, 1)$-distributed.

To show the converse we show the equivalent statement, that if F is not continuous, then $F(X)$ is not $U(0, 1)$-distributed. If F is discontinuous at x, then $0 < \mathrm{P}(X = x) \leq \mathrm{P}(F(X) = F(x))$. Hence, $F(X)$ has a point mass and therefore cannot be $U(0, 1)$-distributed. □

It is not difficult to see that the quantile function, and therefore also VaR, is translation invariant and positive homogeneous. For constants c_1, c_2 with $c_2 > 0$,

$$\begin{aligned} F^{-1}_{c_1+c_2L}(p) &= \min\{m : \mathrm{P}(c_1 + c_2 L \leq m) \geq p\} \\ &= \min\{m : F_L((m - c_1)/c_2) \geq p\} \end{aligned}$$

$$
\begin{aligned}
&= \{\text{put } m' = (m - c_1)/c_2\} \\
&= \min\{c_1 + c_2 m' : F_L(m') \geq p\} \\
&= c_1 + c_2 \min\{m' : F_L(m') \geq p\} \\
&= c_1 + c_2 F_L^{-1}(p). \qquad (6.7)
\end{aligned}
$$

Moreover, the quantile function, and therefore also VaR, satisfies the monotonicity condition. This follows from the fact that $L_2 \leq L_1$ implies $F_{L_1}(m) \leq F_{L_2}(m)$ and therefore

$$
\begin{aligned}
F_{L_1}^{-1}(p) &= \min\{m : F_{L_1}(m) \geq p\} \\
&\geq \min\{m : F_{L_2}(m) \geq p\} = F_{L_2}^{-1}(p).
\end{aligned}
$$

Here we summarize the properties for VaR that have been established up to this point.

Proposition 6.2. *The properties translation invariance, monotonicity, and positive homogeneity hold for* VaR_p.

Examples of the lack of subadditivity of VaR_p can be found even for sums of independent and identically distributed random variables. One such example is obtained by combining Examples 6.9 and 6.10.

Example 6.4 (A crude upper bound). Sometimes we need to estimate the quantile $F_L^{-1}(p)$ for $p \in (0,1)$ close to one, although the distribution of L is far from being well understood. Suppose, for instance, that only the mean $\mathrm{E}[L]$ and the variance $\mathrm{Var}(L)$ are available to us. Cantelli's inequality, the one-sided version of Chebyshev's inequality, says that

$$
\mathrm{P}(L - \mathrm{E}[L] \geq y) \leq \frac{\mathrm{Var}(L)}{y^2 + \mathrm{Var}(L)}
$$

or equivalently that

$$
\mathrm{P}(L \geq y) \leq \frac{\mathrm{Var}(L)}{(y - \mathrm{E}[L])^2 + \mathrm{Var}(L)}.
$$

Now we can turn this upper bound for the tail probability into an upper bound for the p-quantile of the distribution of L:

$$
F_L^{-1}(p) \leq \min\left\{y : \frac{\mathrm{Var}(L)}{(y - \mathrm{E}[L])^2 + \mathrm{Var}(L)} \leq 1 - p\right\} = \mathrm{E}[L] + \left(\frac{\mathrm{Var}(L)p}{1-p}\right)^{1/2}.
$$

The upper bound on the quantile is not necessarily a good estimate, but it is the smallest upper bound (best conservative estimate) if no information about the distribution of L is available besides the mean and the variance.

Example 6.5 (Lognormal distribution). Consider a stock with spot price S_0 today and random spot price S_1 tomorrow, and assume that the 1-day interest rate is zero. We want to compute $\mathrm{VaR}_p(S_1 - S_0)$ under the assumption that the log return $\log(S_1/S_0)$ is normally distributed. Note that $\mathrm{VaR}_p(S_1 - S_0) = F^{-1}_{S_0-S_1}(1-p)$ and that

$$S_0 - S_1 = -S_0(e^{\log(S_1/S_0)} - 1) \stackrel{\mathrm{d}}{=} -S_0(e^{\mu+\sigma Z} - 1),$$

where Z is standard normally distributed. Write $L = S_0 - S_1$ and notice that $L = -g(Z)$, where g is a continuous and strictly increasing function. To compute $\mathrm{VaR}_p(S_1 - S_0) = F_L^{-1}(1-p)$, we will combine the two relations

$$F^{-1}_{-g(Z)}(1-p) = -F^{-1}_{g(Z)}(p), \tag{6.8}$$

$$F^{-1}_{g(Z)}(p) = g(F_Z^{-1}(p)), \tag{6.9}$$

to obtain

$$\mathrm{VaR}_p(S_1 - S_0) = -g(F_Z^{-1}(p)) = S_0(1 - e^{\mu+\sigma\Phi^{-1}(p)}).$$

Let us first show relation (6.8). Since $\mathrm{P}(g(Z) = x) = 0$ for every x, it holds that

$$F_{-g(Z)}(x) = \mathrm{P}(-g(Z) \le x) = \mathrm{P}(g(Z) \ge -x) = 1 - F_{g(Z)}(-x),$$

and therefore solving $F_{-g(Z)}(x) = 1-p$ for x is equivalent to solving $F_{g(Z)}(-x) = p$, which in turn is equivalent to $x = -F^{-1}_{g(Z)}(p)$. Let us now show relation (6.9). We notice that

$$F_{g(Z)}(x) = \mathrm{P}(g(Z) \le x) = \mathrm{P}(Z \le g^{-1}(x)) = F_Z(g^{-1}(x)),$$

and therefore solving $F_{g(Z)}(x) = p$ for x is equivalent to solving $g^{-1}(x) = F_Z^{-1}(p)$, which in turn is equivalent to $x = g(F_Z^{-1}(p))$.

As the previous example illustrates, a common situation is when we want to compute $\mathrm{VaR}_p(X)$ when $X = g(Z)$ for a continuous and monotone function g and a random variable Z. In the preceding example, g was continuous and strictly increasing and Z had a normal distribution. The payoff function of a call option, $(S_1 - K)_+$, is nondecreasing but not strictly increasing, so the preceding calculation does not apply. The following two results show that also the more general situation can be handled without too much difficulty.

Proposition 6.3. *If $g : \mathbb{R} \to \mathbb{R}$ is nondecreasing and left continuous, then for any random variable Z it holds that $F^{-1}_{g(Z)}(p) = g(F^{-1}_Z(p))$ for all $p \in (0, 1)$.*

Proof. First we show that, with $X = g(Z)$, $F^{-1}_X(p) \le g(F^{-1}_Z(p))$. To see this, first observe that since g is nondecreasing, $Z \le F^{-1}_Z(p)$ implies that $g(Z) \le g(F^{-1}_Z(p))$. Moreover, since F_Z is right continuous, it holds that $F_Z(F^{-1}_Z(p)) \ge p$. Therefore,

$$\mathrm{P}(X \le g(F^{-1}_Z(p))) = \mathrm{P}(g(Z) \le g(F^{-1}_Z(p))) \ge \mathrm{P}(Z \le F^{-1}_Z(p)) \ge p.$$

Since $F^{-1}_X(p)$ is the smallest number m such that $\mathrm{P}(X \le m) \ge p$, we have shown that $F^{-1}_X(p) \le g(F^{-1}_Z(p))$.

To show the reverse inequality $F^{-1}_X(p) \ge g(F^{-1}_Z(p))$, we use the left continuity of g. Since g is nondecreasing and left continuous, there exists for each $y \in \mathbb{R}$ and $\varepsilon > 0$ a $\delta > 0$ such that

$$\{z : z \in (y - \delta, y]\} \subset \{z : g(z) \in (g(y) - \varepsilon, g(y)]\}.$$

Moreover, since g is nondecreasing, we have

$$\{z : g(z) \le g(y)\} = \{z : z \le y\} \cup \{z : g(z) = g(y), z > y\},$$
$$\{z : g(z) \le g(y) - \varepsilon\} \subset \{z : z \le y\}.$$

Combining the preceding three set relations yields

$$\{z : g(z) \le g(y) - \varepsilon\} \subset \{z : z \le y - \delta\}.$$

Therefore,

$$\begin{aligned}\mathrm{P}(X \le g(F^{-1}_Z(p)) - \varepsilon) &= \mathrm{P}(g(Z) \le g(F^{-1}_Z(p)) - \varepsilon)\\ &\le \mathrm{P}(Z \le F^{-1}_Z(p) - \delta)\\ &< p,\end{aligned}$$

where in the last step we used the fact that the right continuity of F_Z implies that for every $\delta > 0$ we have $F_Z(F^{-1}_Z(p) - \delta) < p$. It follows that $g(F^{-1}_Z(p)) - \varepsilon < F^{-1}_X(p)$. Since $\varepsilon > 0$ was arbitrary, we conclude that $g(F^{-1}_Z(p)) \le F^{-1}_X(p)$. This completes the proof. □

The following proposition, combined with Proposition 6.3, enables efficient computations of $\mathrm{VaR}_p(X)$ in a wide range of applications. We denote by

$$F^{-1}_X((1 - p)+) = \lim_{\varepsilon \downarrow 0} F^{-1}_X(1 - p + \varepsilon)$$

the limit from the right of the quantile function of X, F^{-1}_X, at the point $1 - p$.

Proposition 6.4. *For any random variable X, $F_{-X}^{-1}(p) = -F_X^{-1}((1-p)+)$ for all $p \in (0,1)$. In particular, if F_X is continuous and strictly increasing, then $F_{-X}^{-1}(p) = -F_X^{-1}(1-p)$.*

The best way to verify the equality $F_{-X}^{-1}(p) = -F_X^{-1}((1-p)+)$ is by selecting a random variable X whose distribution function F_X has both a flat part and a jump and to draw and inspect the graphs of F_X, F_{-X}, F_X^{-1}, and F_{-X}^{-1}. Without loss of generality we may choose a random variable with distribution function $x \mapsto F_X(x)$ whose graph is shown in the upper left plot in Fig. 6.2. To draw the graph of $p \mapsto -F_X^{-1}((1-p)+)$, we proceed as follows. The graph of $p \mapsto F_X^{-1}(p)$ (lower left plot in Fig. 6.2) is obtained by reflecting the graph of F_X in the line $y = x$. Note that the quantile function F_X^{-1} is left continuous. Finally, we draw the graph of $p \mapsto -F_X^{-1}((1-p)+)$ (lower right plot in Fig. 6.2) by first reflecting the graph of F_X^{-1} in x-axis, then reflecting the resulting graph in the line $p = 1/2$, and finally taking limits from the left of the resulting function of p (which corresponds to taking limits from the right if the function is viewed as a function of $1-p$). To draw the graph of $p \mapsto F_{-X}^{-1}(p)$, we proceed as follows. First draw the graph of $x \mapsto F_{-X}(x)$ (upper right plot in Fig. 6.2). Then draw the graph of $p \mapsto F_{-X}^{-1}(p)$ (lower right plot in Fig. 6.2) by reflecting the previous graph in the line $y = x$.

A formal proof of Proposition 6.4 goes as follows.

Proof. First note that

$$\begin{aligned} F_{-X}^{-1}(p) &= \min\{m : \mathrm{P}(-X \le m) \ge p\} \\ &= \min\{m : \mathrm{P}(X \ge -m) \ge p\} \\ &= \min\{m : \mathrm{P}(X < -m) \le 1-p\} \\ &= -\max\{m : \mathrm{P}(X < m) \le 1-p\}. \end{aligned}$$

It remains to show that $\max\{m : \mathrm{P}(X < m) \le 1-p\} = \lim_{\varepsilon\downarrow 0} F_X^{-1}(1-p+\varepsilon)$. Let $m_{1-p} = \max\{m : \mathrm{P}(X < m) \le 1-p\}$, and note that it follows from the definition of the quantile F_X^{-1} that $F_X(F_X^{-1}(u)) = F_X(\min\{m : F_X(m) \ge u\}) \ge u$. Therefore, for $\varepsilon \in (0,p)$,

$$\mathrm{P}(X < m_{1-p}) \le 1-p < 1-p+\varepsilon \le \mathrm{P}(X \le F_X^{-1}(1-p+\varepsilon)),$$

from which it follows that $m_{1-p} \le F_X^{-1}(1-p+\varepsilon)$. Since the inequalities hold for any $\varepsilon \in (0,p)$, we may take the limit as $\varepsilon \downarrow 0$ and therefore conclude that

$$m_{1-p} \le \lim_{\varepsilon\downarrow 0} F_X^{-1}(1-p+\varepsilon).$$

To show the reverse inequality, we first take an arbitrary $\delta > 0$ and note that the definition of m_{1-p} implies that $\mathrm{P}(X < m_{1-p}+\delta) > 1-p$, which in turn implies that

$$\mathrm{P}(X < m_{1-p}+\delta) > 1-p+\varepsilon$$

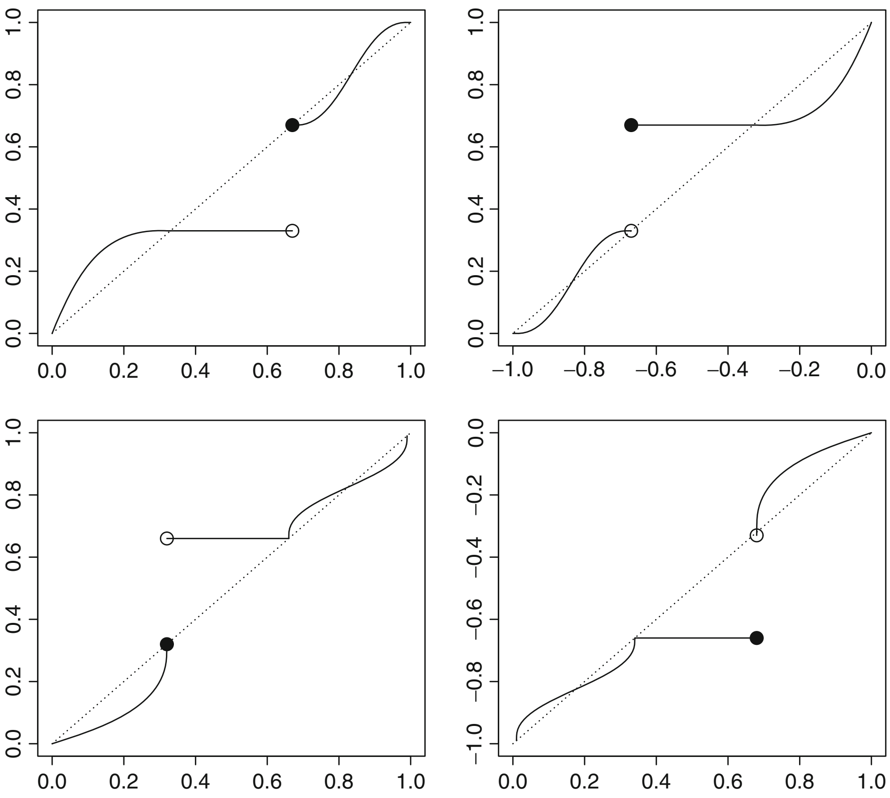

Fig. 6.2 *Upper left plot*: distribution functions $F_X(x)$; *upper right plot*: distribution function $F_{-X}(x)$. *Lower left plot*: quantile function $F_X^{-1}(p)$; *lower right plot*: function $-F_X^{-1}((1-p)+)$

for all sufficiently small $\varepsilon > 0$. Therefore, $m_{1-p} + \delta \geq F_X^{-1}(1-p+\varepsilon)$ for all sufficiently small $\varepsilon > 0$, and taking the limit as $\varepsilon \downarrow 0$ gives

$$m_{1-p} + \delta \geq \lim_{\varepsilon \downarrow 0} F_X^{-1}(1-p+\varepsilon).$$

Since $\delta > 0$ was arbitrary, it follows that

$$m_{1-p} \geq \lim_{\varepsilon \downarrow 0} F_X^{-1}(1-p+\varepsilon).$$

This completes the proof. □

Example 6.6 (Put spread). Consider a portfolio with a value at time 1 given by

$$(K_2 - S_1)_+ - (K_1 - S_1)_+ \quad \text{for } K_1 < K_2.$$

This is the value at maturity of a put spread; a long position in a put option with strike K_2 and price C_2 and a short position of the same size in a put option with a lower strike K_1 and price C_1. The net value X at time 1, considering the cost of the put spread as a risk-free loan to be paid at time 1, is

$$X = (K_2 - S_1)_+ - (K_1 - S_1)_+ - (C_2 - C_1)R_0.$$

We write $L = -X/R_0 = g(S_1)$, where

$$g(y) = \frac{1}{R_0}\Big((K_1 - y)_+ - (K_2 - y)_+\Big) + C_2 - C_1$$

and note that the function g is continuous and nondecreasing. By Proposition 6.3 it follows that

$$\begin{aligned}\mathrm{VaR}_p(X) &= F_L^{-1}(1-p) = F_{g(S_1)}^{-1}(1-p) = g(F_{S_1}^{-1}(1-p)) \\ &= \frac{1}{R_0}\Big((K_1 - F_{S_1}^{-1}(1-p))_+ - (K_2 - F_{S_1}^{-1}(1-p))_+\Big) + C_2 - C_1.\end{aligned}$$

Example 6.7 (Structured product). Financial contracts that are combinations of a bond, giving the buyer a guaranteed return on investment, and a derivative contract, giving the buyer the possibility of a high return, are often called structured products. The simplest form of a structured product with maturity in 1 year is a portfolio consisting of a long position of size h_0 in a risk-free bond that pays 1 to its holder 1 year from now and a long position of size h_1 in a European call option on the value S_1 of a stock index 1 year from now with a strike price K. Suppose that S_1 is lognormally distributed, $\mathrm{LN}(\mu, \sigma^2)$. If the current spot prices of the bond and option are B_0 and C_0, respectively, then the current value of the portfolio is $V_0 = h_0 B_0 + h_1 C_0$.

To evaluate the riskiness of this portfolio, we want to compute $\mathrm{VaR}_p(X)$, where $X = V_1 - V_0/B_0$ and $V_1 = h_0 + h_1(S_1 - K)_+$ is the value of the portfolio at maturity. Write

$$\mathrm{VaR}_p(X) = F_{-B_0 V_1 + V_0}^{-1}(1-p) = F_{-g(Z)}^{-1}(1-p),$$

where Z is standard normally distributed and g is given by

$$g(Z) = B_0(h_0 + h_1(e^{\mu + \sigma Z} - K)_+) - V_0.$$

Note that g is continuous and nondecreasing. Applying first Proposition 6.4 and then Proposition 6.3 gives

$$F_{-g(Z)}^{-1}(1-p) = -F_{g(Z)}^{-1}(p+) = -g(F_Z^{-1}(p+)).$$

Since Z is standard normally distributed with strictly increasing distribution function Φ, we find that $F_Z^{-1}(p+) = \Phi^{-1}(p)$ and conclude that

$$\begin{aligned}\mathrm{VaR}_p(X) &= -B_0(h_0 + h_1(e^{\mu+\sigma\Phi^{-1}(p)} - K)_+) + V_0\\ &= h_1\Big(C_0 - B_0(e^{\mu+\sigma\Phi^{-1}(p)} - K)_+)\Big).\end{aligned}$$

Suppose value X at time 1 of a portfolio can be expressed as $X = f(Z)$ for a smooth nonlinear function f and Z having a standard distribution. If f is not monotone, then it is difficult (or impossible) to express the quantiles of X in terms of the quantiles of Z. One way to overcome this difficulty is by approximating f by a first-order Taylor expansion and then approximate X by

$$X \approx f(\mathrm{E}[Z]) + \frac{df}{dz}(\mathrm{E}[Z])(Z - \mathrm{E}[Z]).$$

This approach is referred to as linearization. The following example gives an illustration of linearization in a simple example where explicit calculations are possible and linearization is not really needed. The example also shows that linearization (like any other approximation) must be used wisely; careless use may result in serious errors.

Example 6.8 (Linearization). You hold a portfolio consisting of a long position of 5 shares of stock A. The stock price today is $S_0 = 100$, and we assume a zero interest rate. The daily log returns

$$Y_1 = \log(S_1/S_0), Y_2 = \log(S_2/S_1), \ldots$$

of stock A are assumed to have a normal distribution with zero mean and standard deviation $\sigma = 0.01$. Let V_0 be the current value of the portfolio, and let $V_1 = S_0 e^{Y_1} = S_0 e^{0.01Z}$, where $Z = Y_1/0.01$ is standard normally distributed, be the value of the portfolio tomorrow.

We first consider the effect of linearization over a 1-day horizon. We start by explicitly computing $\mathrm{VaR}_{0.01}(V_1 - V_0)$ and then compute the approximation obtained by replacing V_1 by its first-order Taylor approximation with respect to Z. Notice that $\mathrm{VaR}_{0.01}(V_1 - V_0) = F^{-1}_{V_0-V_1}(0.99)$ and $V_0 - V_1 = -500(e^{Y_1} - 1) = -500(e^{0.01Z} - 1)$. Therefore, as in Example 6.5,

$$\mathrm{VaR}_{0.01}(V_1 - V_0) = 500(1 - e^{0.01\Phi^{-1}(0.01)}) = 11.5.$$

The first-order Taylor approximation of V_1 is $V_1 = 500e^{0.01Z} \approx 500(0.01Z + 1)$, which gives

$$\mathrm{VaR}_{0.01}(V_1 - V_0) \approx \mathrm{VaR}_{0.01}(5Z) = 5\Phi^{-1}(0.99) \approx 11.6.$$

The relative error of the VaR approximation is 1.2%, which is rather small.

We now consider the effect of linearization over a longer time horizon and illustrate that the error due to linearization may be substantial. We make the simplifying assumption that log returns over nonoverlapping time periods are independent. We consider the effect of holding the aforementioned portfolio for 100 (trading) days and let V_{100} be the value of the portfolio 100 days from now. We start by explicitly computing $\mathrm{VaR}_{0.01}(V_{100} - V_0)$ and then compute the approximation obtained by replacing V_{100} seen as a function of a standard normal variable Z by its first-order Taylor approximation with respect to Z. As previously, we ignore interest rates. We may write $\mathrm{VaR}_{0.01}(V_{100}-V_0) = F^{-1}_{V_0-V_{100}}(0.99)$, where $V_0-V_{100} = -500(e^{Y_{100}} - 1)$ with Y_{100} denoting the 100-day log return. Note that

$$Y_{100} = \log S_{100}/S_0 = \log S_1/S_0 + \cdots + \log S_{100}/S_{99},$$

which shows that Y_{100} is a sum of 100 independent $\mathrm{N}(0, 0.01^2)$-distributed random variables. Therefore, $Y_{100} \stackrel{\mathrm{d}}{=} 0.1Z$, where Z is standard normally distributed. In particular, $V_0 - V_{100} \stackrel{\mathrm{d}}{=} -500(e^{0.1Z} - 1)$ and

$$\mathrm{VaR}_{0.01}(V_{100} - V_0) = 500(1 - e^{0.1\Phi^{-1}(0.01)}) \approx 103.8.$$

Using a first-order Taylor approximation gives $V_{100} \stackrel{\mathrm{d}}{=} 500e^{0.1Z} \approx 500(0.1Z + 1)$, which gives the approximation

$$\mathrm{VaR}_{0.01}(V_{100} - V_0) \approx \mathrm{VaR}_{0.01}(50Z) = 50\Phi^{-1}(0.99) \approx 116.3.$$

The relative error of the VaR approximation is 12.1%.

Next follows the first two in a series of four examples on credit default swaps (CDSs). The examples treat portfolios containing defaultable bonds and CDSs. There are two general messages communicated by these examples. The first message is that VaR at level p does not provide any information about the worst-case outcomes corresponding to an event whose probability is less than p. Using VaR, therefore, enables investors to hide risk in the right tail of the distribution of L. The second message is that, besides not investing at all, there are essentially two ways to reduce risk. One way is to hedge a risk by buying protection against undesired events. Hedging may be viewed as buying insurance. In the following examples, hedging amounts to buying credit default swaps. Another way to reduce risk is by diversification. A well-diversified position has a future value that depends on many independent sources of randomness such that the exposure to each one of them is small. Diversification is the key principle for an insurer and the opposite of buying insurance.

A problem with VaR is that it does not necessarily reward diversification. In Example 6.10 it is shown that a diversified portfolio may have higher risk, measured by VaR, than a comparable nondiversified portfolio. The example also shows that VaR is not subadditive in general.

Example 6.9 (Credit default swap I). Consider an investor with \$100 who has the opportunity to take long positions in a defaultable bond and a credit default swap (CDS) on this bond. One bond costs \$97 now and pays \$100 6 months from now if the issuer does not default and 0 if the issuer defaults. The CDS costs \$4 and pays \$100 6 months from now if the bond issuer defaults and nothing otherwise. For simplicity we assume that a risk-free bond with maturity in 6 months has zero interest rate, so $B_0 = 1$. The investor believes that the default probability is 0.02 and wants to maximize the expected value of V_1, the value in dollars of the investor's position at the maturity of the bond, subject to the risk constraint $\mathrm{VaR}_{0.05}(V_1 - 100) \leq 10$ and a budget constraint. It is assumed throughout that the investor can only take long positions. Otherwise, with the prices given previously, there would be an arbitrage opportunity. Why? How much of the \$100 does the investor invest in the bond? How much in the CDS?

Let w_1 and w_2 be the amounts invested in bonds and CDSs in the portfolio, respectively. Let $c_1 = 97$ and $c_2 = 4$ be the prices of the bond and the CDS, respectively. Then the value at time 1 (after 6 months) is $V_1 = w_1 c_1^{-1} 100(1 - I) + w_2 c_2^{-1} 100 I$, where I is the default indicator, $I = 1$ if the issuer defaults, and $I = 0$ otherwise, with $\mathrm{P}(I = 1) = 0.02$. Then

$$\mathrm{E}[V_1] = 98 w_1 c_1^{-1} + 2 w_2 c_2^{-1} = \frac{98}{97} w_1 + \frac{1}{2} w_2,$$

from which it is clear that the investor wants to invest as much as possible in the bond without violating the constraints. Moreover,

$$\mathrm{VaR}_p(V_1 - 100) = 100 + \mathrm{VaR}_p\left(100 w_1 c_1^{-1} + 100\left(w_2 c_2^{-1} - w_1 c_1^{-1}\right) I\right),$$

which gives

$$\begin{aligned}\mathrm{VaR}_p(V_1 - 100) =& 100 - 100 w_1 c_1^{-1} \\ &+ \begin{cases} 100(w_2 c_2^{-1} - w_1 c_1^{-1})\,\mathrm{VaR}_p(I) & \text{if } w_2 c_2^{-1} \geq w_1 c_1^{-1}, \\ 100(w_1 c_1^{-1} - w_2 c_2^{-1})\,\mathrm{VaR}_p(-I) & \text{if } w_2 c_2^{-1} < w_1 c_1^{-1}. \end{cases}\end{aligned}$$

By (6.6) we have $\mathrm{VaR}_p(I) = F_{-I}^{-1}(1 - p)$ and $\mathrm{VaR}_p(-I) = F_I^{-1}(1 - p)$ where, by Proposition 6.4,

$$F_{-I}^{-1}(1 - p) = -F_I^{-1}(p+) = \begin{cases} -1 & \text{if } p \in [0.98, 1], \\ 0 & \text{if } p \in [0, 0.98), \end{cases}$$

and

$$F_I^{-1}(1-p) = \begin{cases} 0 \text{ if } p \in [0.02, 1], \\ 1 \text{ if } p \in [0, 0.02). \end{cases}$$

This implies that

$$\mathrm{VaR}_p(V_1 - 100) = 100 - \begin{cases} 100 \max(w_1 c_1^{-1}, w_2 c_2^{-1}) & \text{if } p \in [0.98, 1], \\ 100 w_1 c_1^{-1} & \text{if } p \in [0.02, 0.98), \\ 100 \min(w_1 c_1^{-1}, w_2 c_2^{-1}) & \text{if } p \in [0, 0.02). \end{cases}$$

In particular, $\mathrm{VaR}_{0.05}(V_1 - 100) = 100 - 100 w_1 c_1^{-1}$, and therefore $w_2 = 100 - w_1$, together with $\mathrm{VaR}_{0.05}(V_1 - 100) \le 10$, is equivalent to $w_1 \ge 87.3$. Since a dollar invested in the bond gives a much better expected return than a dollar invested in the CDS, the investor wants to maximize w_1 subject to the constraints. Therefore, the solution to the optimization problem with the VaR constraint is $(w_1, w_2) = (100, 0)$. That is, buy defaultable bonds only. The catch here is that VaR at level 0.05 does not take into account the possibility of default, which occurs with probability 0.02. This enables the investor to hide the default risk in the tail.

Example 6.10 (Credit default swap II). Let us look a bit closer at the optimal solution $(w_1, w_2) = (100, 0)$ to the investment problem in Example 6.9. The optimal weights give the optimal portfolio value $V_1 = (100^2/97)(1 - I)$ at maturity. Moreover, we have seen that

$$\begin{aligned} \mathrm{VaR}_{0.05}(V_1 - 100) &= \mathrm{VaR}_{0.05}\left(\frac{100^2}{97}(1 - I) - 100\right) \\ &= 100 - \frac{100^2}{97} + \frac{100^2}{97}\,\mathrm{VaR}_{0.05}(-I) \\ &= 100\Big(1 - \frac{100}{97}\Big) < 0. \end{aligned} \tag{6.10}$$

The negative value highlights the fact that at the 5% level VaR does not pick up the default risk. In particular, it treats the defaultable bond as a risk-free bond. Suppose, in contrast, that we have 100 identical bonds whose default events are independent and that the investor invests one dollar in each of them (which gives the same expected portfolio value as for the optimal solution in Example 6.9). The risk of the new portfolio, in terms of $\mathrm{VaR}_{0.05}$, is

$$\begin{aligned} \mathrm{VaR}_{0.05}(V_1 - 100) &= \mathrm{VaR}_{0.05}\left(\frac{100}{97}\sum_{k=1}^{100}(1 - I_k) - 100\right) \\ &= 100 - \frac{100^2}{97} + \frac{100}{97}\,\mathrm{VaR}_{0.05}\left(-\sum_{k=1}^{100} I_k\right). \end{aligned}$$

Since $Z = \sum_{k=1}^{100} I_k$ is Bin(100, 0.02)-distributed and $\mathrm{VaR}_{0.05}(-Z) = F_Z^{-1}(0.95)$, it follows that

$$\mathrm{VaR}_{0.05}(V_1 - 100) = 100 - \frac{100^2}{97} + \frac{100}{97} F_Z^{-1}(0.95).$$

We can compute $\mathrm{P}(Z \leq 4) \approx 0.949$ and $\mathrm{P}(Z \leq 5) \approx 0.985$. Therefore, $F_Z^{-1}(0.95) = \min\{m : \mathrm{P}(Z \leq m) \geq 0.95\} = 5$, which implies that

$$\begin{aligned}\mathrm{VaR}_{0.05}(V_1 - 100) &= 100 + \frac{100}{97}(-100 + 5) \\ &= 100\Big(1 - \frac{95}{97}\Big) > 0. \qquad (6.11)\end{aligned}$$

That is, in this example diversification increases the risk! The reason is that diversification here makes $\mathrm{VaR}_{0.05}$ take into account the default risk that for the nondiversified investment was hidden in the tail. In particular, we conclude that VaR is not subadditive since (6.10) and (6.11) imply

$$\mathrm{VaR}_{0.05}\left(\sum_{k=1}^{100}(1 - I_k)\right) > \sum_{k=1}^{100} \mathrm{VaR}_{0.05}(1 - I_k).$$

6.3 Expected Shortfall

Although VaR is probably the most commonly used risk measure for risk control in the financial industry, it has several limitations. Its biggest weakness is that it ignores the left tail (beyond level p) of the distribution of X. (The fact that it is just a quantile value means that it ignores most of the distribution of X.) In particular, it allows a careless/dishonest risk manager to miss/hide unlikely but catastrophic risks in the left tail.

A natural remedy for not considering catastrophic loss events with small probabilities would be to consider the average VaR values below the level p. This average of VaR values gives the risk measure expected shortfall (ES) at level p, which is defined as

$$\mathrm{ES}_p(X) = \frac{1}{p}\int_0^p \mathrm{VaR}_u(X)\,du.$$

With minor technical modifications, ES is also called Average VaR (AVaR), Conditional VaR (CVaR), Tail VaR (TVaR), or Tail Conditional Expectation (TCE).

ES is often proposed as a superior alternative to VaR because it considers all of the left tail of the probability distribution of X and because it is a coherent measure of risk. The coherence of ES, Proposition 6.6, implies that it is also convex, and the latter property is essential for ensuring that investment problems with ES constraints are convex optimization problems. To show the coherence of ES and also to use it effectively in optimization problems, we first present useful alternative representations of ES.

Proposition 6.5. *(i) ES has the following representations:*

$$\mathrm{ES}_p(X) = \frac{1}{p}\int_{1-p}^{1} F_L^{-1}(u)du, \quad L = -X/R_0, \tag{6.12}$$

$$\mathrm{ES}_p(X) = -\frac{1}{p}\int_0^p F_{X/R_0}^{-1}(u)du, \tag{6.13}$$

$$\begin{aligned}\mathrm{ES}_p(X) = &-\frac{1}{p}\,\mathrm{E}[X/R_0 I\{X/R_0 \le F_{X/R_0}^{-1}(p)]\\ &- F_{X/R_0}^{-1}(p)\left(1 - \frac{F_{X/R_0}(F_{X/R_0}^{-1}(p))}{p}\right),\end{aligned} \tag{6.14}$$

$$\mathrm{ES}_p(X) = \min_c \; -c + \frac{1}{p}\,\mathrm{E}[(c - X/R_0)_+]. \tag{6.15}$$

(ii) If X has a continuous distribution function, then, with $L = -X/R_0$,

$$\mathrm{ES}_p(X) = \mathrm{E}[L \mid L \ge \mathrm{VaR}_p(X)] = \mathrm{E}[L \mid L \ge F_L^{-1}(1-p)]. \tag{6.16}$$

The right-hand side of (6.15) is often called CVaR. This representation is useful in portfolio optimization problems. From (6.16) we find that if X has a continuous distribution function, then ES is the average loss conditional on the loss being larger than or equal to the VaR at the level p. This expression motivates the name ES.

Proof. (i) From the definition we see that ES is simply an average of quantile values of L:

$$\mathrm{ES}_p(X) = \frac{1}{p}\int_0^p F_L^{-1}(1-u)du = \frac{1}{p}\int_{1-p}^{1} F_L^{-1}(u)du.$$

This proves the first representation (6.12). To prove the second representation, recall from Proposition 6.4 that $F_{-X/R_0}^{-1}(1-u) = -F_{X/R_0}^{-1}(u+)$. But $F_{X/R_0}^{-1}(u+)$ is not equal to $F_{X/R_0}^{-1}(u)$ in general. However, we do have equality for almost all u in the sense that if we draw U uniformly on $(0,1)$, then $F_{X/R_0}^{-1}(U+) = F_{X/R_0}^{-1}(U)$ with probability one. In particular, when U has a uniform distribution on $(0,p)$, it holds that

$$\mathrm{ES}_p(X) = \frac{1}{p}\int_0^p F_L^{-1}(1-u)du$$

$$
\begin{aligned}
&= \mathrm{E}[F_L^{-1}(1-U)] \\
&= -\mathrm{E}[F_{X/R_0}^{-1}(U+)] \\
&= -\mathrm{E}[F_{X/R_0}^{-1}(U)] \\
&= -\frac{1}{p}\int_0^p F_{X/R_0}^{-1}(u)du,
\end{aligned}
$$

which proves (6.13). Let us prove (6.14). The only difficulty is when F_{X/R_0} has a jump at $F_{X/R_0}^{-1}(p)$ and $F_{X/R_0}(F_{X/R_0}^{-1}(p)) > p$. Using statements (i) and (iii) of Proposition 6.1 shows that

$$
\begin{aligned}
&\mathrm{E}[X/R_0 I\{X/R_0 \le F_{X/R_0}^{-1}(p)\}] \\
&\quad = \mathrm{E}[F_{X/R_0}^{-1}(U) I\{F_{X/R_0}^{-1}(U) \le F_{X/R_0}^{-1}(p)] \\
&\quad = \mathrm{E}[F_{X/R_0}^{-1}(U) I\{U \le F_{X/R_0}(F_{X/R_0}^{-1}(p))] \\
&\quad = \mathrm{E}[F_{X/R_0}^{-1}(U) I\{U \le p\}] + \mathrm{E}[F_{X/R_0}^{-1}(U) I\{p < U \le F_{X/R_0}(F_{X/R_0}^{-1}(p))\}] \\
&\quad = \int_0^p F_{X/R_0}^{-1}(u)du + F_{X/R_0}^{-1}(p)\Big(F_{X/R_0}(F_{X/R_0}^{-1}(p)) - p\Big).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&-\frac{1}{p}\mathrm{E}[X/R_0 I\{X/R_0 \le F_{X/R_0}^{-1}(p)] - F_{X/R_0}^{-1}(p)\left(1 - \frac{F_{X/R_0}(F_{X/R_0}^{-1}(p))}{p}\right) \\
&\quad = -\frac{1}{p}\int_0^p F_{X/R_0}^{-1}(u)du,
\end{aligned}
$$

from which the conclusion follows from (6.13). To prove (6.15) we consider the function

$$
G(c) = -c + \frac{1}{p}\mathrm{E}[(c - X/R_0)_+] = -c + \frac{1}{p}\int_{-\infty}^c F_{X/R_0}(x)dx
$$

and note that G is convex. Since

$$
\begin{aligned}
\mathrm{E}[(c - X/R_0)_+] &= \mathrm{E}[(c - X/R_0) I\{c - X/R_0 > 0\}] \\
&= \int_0^\infty \mathrm{P}(c - X/R_0 > t)dt \\
&= \int_{-\infty}^c \mathrm{P}(X/R_0 < u)du,
\end{aligned}
$$

we find that G is differentiable except at the points where F_{X/R_0} has jumps and, except for those points,

$$G'(c) = -1 + \frac{1}{p} F_{X/R_0}(c).$$

It follows that $G'(c) \leq 0$ for c such that $F_{X/R_0}(c) \leq p$ and that $G'(c) \geq 0$ for c such that $F_{X/R_0}(c) \geq p$. Therefore, G has a (not necessarily unique) minimum at $\min\{c : F_{X/R_0}(c) \geq p\} = F^{-1}_{X/R_0}(p)$. Evaluating G at this point gives

$$\begin{aligned}
G(F^{-1}_{X/R_0}(p)) &= -F^{-1}_{X/R_0}(p) + \frac{1}{p}\,\mathrm{E}[(F^{-1}_{X/R_0}(p) - X/R_0)_+] \\
&= -F^{-1}_{X/R_0}(p) + \frac{1}{p}\,\mathrm{E}\left[\left(F^{-1}_{X/R_0}(p) - X/R_0\right) I\left\{X/R_0 \leq F^{-1}_{X/R_0}(p)\right\}\right] \\
&= -\frac{1}{p}\,\mathrm{E}\left[X/R_0 I\left\{X/R_0 \leq F^{-1}_{X/R_0}(p)\right\}\right] \\
&\quad - F^{-1}_{X/R_0}(p)\left(1 - \frac{F_{X/R_0}(F^{-1}_{X/R_0}(p))}{p}\right),
\end{aligned}$$

from which the conclusion follows from (6.14).

(ii) Suppose that X has a continuous distribution function. Recall from point (iii) of Proposition 6.1 that if U is uniformly distributed on $(0,1)$, then $F_L^{-1}(U)$ has distribution function F_L. In particular, the random variables L, $F_L^{-1}(U)$, and $F_L^{-1}(1-U)$ all have the same distribution function F_L. Moreover, if F_L is continuous, then $F_L(F_L^{-1}(u)) = u$ by Proposition 6.1(ii). Therefore,

$$\begin{aligned}
\mathrm{E}[L \mid L \geq F_L^{-1}(1-p)] &= \frac{\mathrm{E}[L\, I\{L \geq F_L^{-1}(1-p)\}]}{\mathrm{P}(L \geq F_L^{-1}(1-p))} \\
&= \frac{1}{p}\,\mathrm{E}[F_L^{-1}(1-U)\, I\{F_L^{-1}(1-U) \geq F_L^{-1}(1-p)\}] \\
&= \frac{1}{p}\,\mathrm{E}[F_L^{-1}(1-U)\, I\{1-U \geq 1-p\}] \\
&= \frac{1}{p}\,\mathrm{E}[F_L^{-1}(1-U)\, I\{U \leq p\}] \\
&= \frac{1}{p}\int_0^p \mathrm{VaR}_u(X)\,du.
\end{aligned}$$

□

We are now well equipped to prove that ES is a coherent measure of risk.

Proposition 6.6. *ES is a coherent measure of risk.*

Proof. It follows immediately from the definition that ES inherits the properties translation invariance, monotonicity, and positive homogeneity from VaR. It only remains to prove subadditivity. Consider two future portfolio values X_1 and X_2 and write $Y_k = X_k/R_0$ for $k = 1, 2$. We will use representation (6.15) of ES to prove subadditivity, i.e., that $\mathrm{ES}_p(X_1 + X_2) \leq \mathrm{ES}_p(X_1) + \mathrm{ES}_p(X_2)$. For $k = 1, 2$ let c_k^* be a minimizer of

$$-c + \frac{1}{p}\,\mathrm{E}[(c - Y_k)_+].$$

Note that

$$\begin{aligned}\mathrm{ES}_p(X_1 + X_2) &= \min_c -c + \frac{1}{p}\,\mathrm{E}[(c - Y_1 - Y_2)_+] \\ &\leq -(c_1^* + c_2^*) + \frac{1}{p}\,\mathrm{E}\left[(c_1^* + c_2^* - Y_1 - Y_2)_+\right].\end{aligned}$$

The proof is complete if we show the nonnegativity of the difference

$$\begin{aligned}&\mathrm{ES}_p(X_1) + \mathrm{ES}_p(X_2) - \mathrm{ES}_p(X_1 + X_2) \\ &\quad\geq -c_1^* + \frac{1}{p}\,\mathrm{E}[(c_1^* - Y_1)_+] - c_2^* + \frac{1}{p}\,\mathrm{E}[(c_2^* - Y_2)_+] \\ &\qquad + (c_1^* + c_2^*) - \frac{1}{p}\,\mathrm{E}[(c_1^* + c_2^* - Y_1 - Y_2)_+] \\ &\quad= \frac{1}{p}\,\mathrm{E}[(c_1^* - Y_1)(I\{Y_1 \leq c_1^*\} - I\{Y_1 + Y_2 \leq c_1^* + c_2^*\})] \\ &\qquad + \frac{1}{p}\,\mathrm{E}[(c_2^* - Y_2)(I\{Y_2 \leq c_2^*\} - I\{Y_1 + Y_2 \leq c_1^* + c_2^*\})].\end{aligned}$$

We claim that the last two terms above are nonnegative. Indeed,

$$\begin{aligned}&\mathrm{E}[(c_1^* - Y_1)(I\{Y_1 \leq c_1^*\} - I\{Y_1 + Y_2 \leq c_1^* + c_2^*\})] \\ &\quad= \mathrm{E}[(c_1^* - Y_1)(I\{Y_1 \leq c_1^*\} - I\{Y_1 + Y_2 \leq c_1^* + c_2^*\})I\{Y_1 \leq c_1^*\}] \\ &\qquad + \mathrm{E}[(c_1^* - Y_1)(I\{Y_1 \leq c_1^*\} - I\{Y_1 + Y_2 \leq c_1^* + c_2^*\})I\{Y_1 > c_1^*\}] \\ &\quad\geq \mathrm{E}[(c_1^* - Y_1)I\{Y_1 \leq c_1^*\}] - \mathrm{E}[(c_1^* - Y_1)I\{Y_1 > c_1^*\}] \\ &\quad\geq 0,\end{aligned}$$

which shows the nonnegativity of the first term. An identical argument shows that the second term is nonnegative too. The proof is complete. □

Next we continue the sequence of examples on defaultable bonds and CDSs. Here the risk measure VaR is replaced by ES, and this changes the portfolio selection problem substantially.

Example 6.11 (Credit default swap III). Consider the investor and the investment opportunities in Example 6.9. Here the risk constraint $\mathrm{VaR}_{0.05}(V_1 - 100) \le 10$ is replaced by $\mathrm{ES}_{0.05}(V_1 - 100) \le 10$.

Recall that $\mathrm{VaR}_p(V_1 - 100)$ was computed in Example 6.9:

$$\mathrm{VaR}_p(V_1 - 100) = 100 - \begin{cases} 100\max(w_1 c_1^{-1}, w_2 c_2^{-1}) & \text{if } p \in [0.98, 1], \\ 100 w_1 c_1^{-1} & \text{if } p \in [0.02, 0.98), \\ 100\min(w_1 c_1^{-1}, w_2 c_2^{-1}) & \text{if } p \in [0, 0.02). \end{cases}$$

Then $\mathrm{ES}_{0.05}(V_1 - 100)$ can be computed as

$$\begin{aligned} \mathrm{ES}_{0.05}(V_1 - 100) &= \frac{1}{0.05}\int_0^{0.05} \mathrm{VaR}_p(V_1 - 100)dp \\ &= \begin{cases} 100 - 100 w_1 c_1^{-1} & \text{if } w_1 c_1^{-1} < w_2 c_2^{-1}, \\ 100 - 100\frac{3}{5} w_1 c_1^{-1} - 100\frac{2}{5} w_2 c_2^{-1} & \text{if } w_1 c_1^{-1} \ge w_2 c_2^{-1}. \end{cases} \end{aligned}$$

Recall that $c_1 = 97$ and $c_2 = 4$. With $w_2 = 100 - w_1$ we find that $w_1 c_1^{-1} < w_2 c_2^{-1}$ is equivalent to $w_1 < 96.0396$. We want to take w_1 as large as possible and therefore consider the case $w_1 \ge 96.0396$. In this case, $\mathrm{ES}_{0.05}(V_1 - 100) \le 10$, together with $w_2 = 100 - w_1$, is equivalent to $w_1 \le 97$. Since a dollar invested in the bond gives a much better expected return than a dollar invested in the CDS, the investor wants to maximize w_1 subject to the constraints. Therefore, the solution to the optimization problem with the ES constraint is $(w_1, w_2) = (97, 3)$. Since ES takes into account the entire tail, there is no way to hide the default risk in the tail. This is reflected in the optimal portfolio.

Example 6.12 (Credit default swap IV). Consider an investor who has \$100 and may invest the capital in long positions in 100 bonds and CDSs that are identical to those in Example 6.10. It is assumed that the corresponding indicator variables I_k (I_k takes the value 1 if the kth bond issuer defaults) are independent. The value of the investor's portfolio at the maturity of the bonds is

$$V_1 = \sum_{k=1}^{100} \frac{100}{97} w_k (1 - I_k) + \sum_{k=1}^{100} \frac{100}{4} w_{100+k} I_k,$$

where $w_1, \dots, w_{100}$ is the capital invested in the bonds and $w_{101}, \dots, w_{200}$ is the capital invested in the CDSs. The investor wants to maximize the expected value

$$\mathrm{E}[V_1] = \sum_{k=1}^{100} \frac{98}{97} w_k + \sum_{k=1}^{100} \frac{1}{2} w_{100+k},$$

from which it is seen that the investor wants to invest as much as possible in the bonds. The risk constraint is given by $\mathrm{ES}_{0.05}(V_1 - 100) \leq 10$. In Example 6.11, we saw that with only one bond and one CDS the optimal solution was $(w_1, w_2) = (97, 3)$. Here it seems plausible that a diversified position in the bonds leads to lower risk and therefore that it will be possible to invest less capital in the CDSs with the low expected returns. We now verify that this is indeed the case. Just as in Example 6.10 we have

$$\mathrm{VaR}_p(V_1 - 100) = 100 - \frac{100^2}{97} + \frac{100}{97} F_Z^{-1}(1-p),$$

where $Z = \sum_{k=1}^{100} I_k$ is Bin(100, 0.02)-distributed. This gives

$$\mathrm{ES}_p(V_1 - 100) = 100 - \frac{100^2}{97} + \frac{100}{97}\left(\frac{1}{0.05}\int_0^{0.05} F_Z^{-1}(1-p)dp\right),$$

where

$$\begin{aligned}\frac{1}{0.05}\int_0^{0.05} F_Z^{-1}(1-p)dp &= 20\Big((\mathrm{P}(Z \leq 5) - 0.95)5 \\ &\quad + \sum_{k=6}^{100} k(\mathrm{P}(Z \leq k) - \mathrm{P}(Z \leq k-1))\Big) \\ &\approx 5.41416,\end{aligned}$$

and therefore $\mathrm{ES}_{0.05}(V_1 - 100) \approx 2.488825 < 10$. We conclude that the investor may invest the entire capital in the bonds without violating the risk constraint. Thus, an optimal portfolio is $w_1 = \cdots = w_{100} = 1, w_{101} = \cdots = w_{200} = 0$.

Next we study some standard models for log returns of asset prices where ES can be explicitly computed.

Example 6.13 (Normal and Student's t distribution). Consider a 1-day investment in a risky asset. Suppose the influence of interest rates for such a short time period can be neglected. Let $X = V_1 - V_0 = \mu + \sigma Z$, where Z is a standard normally distributed random variable, and let Φ and ϕ denote the distribution and density function of Z, respectively. Then $\mathrm{VaR}_p(X) = -\mu + \sigma\Phi^{-1}(1-p)$ and

$$\begin{aligned}\mathrm{ES}_p(X) &= -\mu + \frac{\sigma}{p}\int_{1-p}^{1} \Phi^{-1}(u)du \\ &= \{\text{set } l = \Phi^{-1}(u)\} \\ &= -\mu + \frac{\sigma}{p}\int_{\Phi^{-1}(1-p)}^{\infty} l\phi(l)dl\end{aligned}$$

$$= -\mu + \frac{\sigma}{p} \int_{\Phi^{-1}(1-p)}^{\infty} l \frac{1}{\sqrt{2\pi}} e^{-l^2/2} \mathrm{d}l$$
$$= -\mu + \frac{\sigma}{p} \Big[- \frac{1}{\sqrt{2\pi}} e^{-l^2/2} \Big]_{\Phi^{-1}(1-p)}^{\infty}$$
$$= -\mu + \sigma \frac{\phi(\Phi^{-1}(1-p))}{p}.$$

Now let Z have a standard Student's t distribution with $\nu > 0$ degrees of freedom. Then Z has a density

$$g_\nu(x) = C \Big(1 + \frac{x^2}{\nu}\Big)^{-(\nu+1)/2}, \quad \text{where } C = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}.$$

If t_ν is the distribution function of Z, then $\mathrm{VaR}_p(X) = -\mu + \sigma t_\nu^{-1}(1-p)$ and, if $\nu > 1$, then

$$\mathrm{ES}_p(X) = -\mu + \frac{\sigma}{p} \int_{t_\nu^{-1}(1-p)}^{\infty} l g_\nu(l) dl$$
$$= -\mu + \frac{\sigma}{p} \left[\frac{C\nu/2}{-(\nu+1)/2+1} \Big(1 + \frac{l^2}{\nu}\Big)^{-(\nu+1)/2+1} \right]_{t_\nu^{-1}(1-p)}^{\infty}$$
$$= -\mu + \sigma \frac{g_\nu(t_\nu^{-1}(1-p))}{p} \Big(\frac{\nu + (t_\nu^{-1}(p))^2}{\nu - 1}\Big).$$

Example 6.14 (Normal and Student's t: a comparison). The normal distribution and the Student's t distribution are simple and popular distributions for modeling log returns of asset prices. An important difference between the two is that Student's t distributions have heavier tails, i.e., they place more mass far away from the mean. This can be observed directly from the density function. The standard Student's t distribution with ν degrees of freedom has a density that decays roughly as $|x|^{-\nu}$ for large $|x|$ (called polynomial decay), whereas the standard normal density decays much faster, as $e^{-x^2/2}$. This implication of heavy tails is that there is a higher probability of extreme outcomes. Let us compare the risk measures VaR and ES for the two distributions.

First we compare $\mathrm{VaR}_p(X)$ and $\mathrm{ES}_p(X)$ as a function of p (left plot in Fig. 6.3). The plot shows the ratio $\mathrm{ES}_p(X)/\,\mathrm{VaR}_p(X)$ as a function of p for the standard normal distribution (lower graph) and the standard Student's t distribution with 3, 2, and 1.1 degrees of freedom (second lowest to upper graph). For the normal distribution the ratio is slightly above one, indicating that for small values of p most of the remaining probability mass in the tail to the left of $\Phi^{-1}(p)$ is concentrated very close to $\Phi^{-1}(p)$. Note that for heavier tails, i.e., smaller degree of freedom

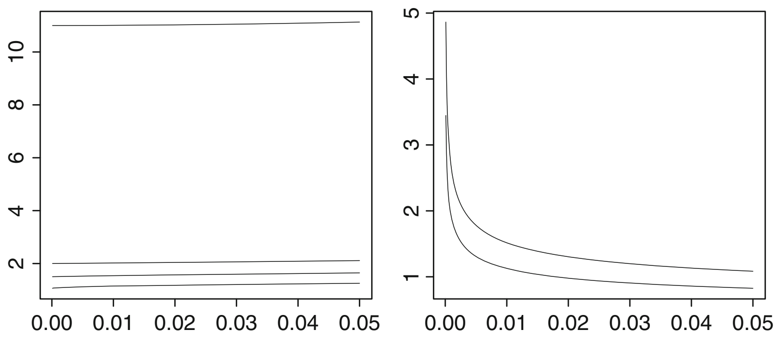

Fig. 6.3 *Left plot*: graphs of $\mathrm{ES}_p(X)/\,\mathrm{VaR}_p(X)$ as a function of p for X standard normal distribution (*lowest graph*) and Student's t-distribution for $\nu = 3, 2, 1.1$. *Right plot*: graphs of $\mathrm{VaR}_p(X)/\,\mathrm{VaR}_p(Y)$ (*lower graph*) and $\mathrm{ES}_p(X)/\,\mathrm{ES}_p(Y)$ (*upper graph*) as functions of p, where X is t-distributed with $\nu = 3$ and variance 1, and Y is standard normally distributed

parameters ν, the ratio is higher, indicating that the probability mass to the left of $t_\nu^{-1}(p)$ is spread out to the left of this value and spread out more the smaller the value of ν is.

In the right plot in Fig. 6.3, we compare VaR for a t_3-distribution and a normal distribution with unit variance, and similarly for ES by plotting the ratios $\mathrm{VaR}_p(X)/\,\mathrm{VaR}_p(Y)$ and $\mathrm{ES}_p(X)/\,\mathrm{ES}_p(Y)$ as functions of p, where X is t-distributed with $\nu = 3$ and variance 1, and Y is standard normally distributed. If Z has a standard t_ν-distribution, then its variance is $\nu(\nu-2)^{-1}$, so in this example, $X = Z/\sqrt{3}$, which implies that X and Y both have unit variance. We observe that for small p the ratios are greater than one. This is a result of the heavier tails of the t_3-distribution.

Example 6.15 (Lognormal distribution). Consider the current and future values V_0 and V_1 of an asset. By borrowing the amount V_0 to finance the long position in the asset, the future net value of the position is $X = V_1 - V_0 R_0$, where $V_0 R_0$ is the future value of the debt. If $Z_1 = \log(V_1/V_0)$ is the log return of the asset, then $X = V_0(\exp\{Z_1\} - R_0)$.

We will analyze $\mathrm{ES}_p(X)$ under the assumption that Z_1 has either a normal distribution or a Student's t distribution. Applying Proposition 6.3, with $g(z) = V_0(e^z - R_0)$, and Proposition 6.4 gives

$$\mathrm{VaR}_u(X) = F^{-1}_{-g(Z_1)/R_0}(1-u) = -g\Big(F^{-1}_{Z_1}(u)\Big) = V_0\left(1 - \frac{1}{R_0} e^{F^{-1}_{Z_1}(u)}\right).$$

If Z_1 is $\mathrm{N}(\mu, \sigma^2)$-distributed, then $F^{-1}_{Z_1}(u) = \mu + \sigma\Phi^{-1}(u)$

$$\mathrm{ES}_p(X) = \frac{1}{p}\int_0^p \mathrm{VaR}_u(X)du = V_0\left(1 - \frac{1}{pR_0}\int_0^p e^{\mu+\sigma\Phi^{-1}(u)}du\right).$$

With $q(u) = \Phi^{-1}(u)$ we have $dq(u)/du = 1/\phi(\Phi^{-1}(u))$, and the integral to the right above can be written as

$$\begin{aligned}\int_0^p e^{\mu+\sigma\Phi^{-1}(u)}du &= \int_{-\infty}^{\Phi^{-1}(p)} \frac{1}{\sqrt{2\pi}} e^{\mu+\sigma q-q^2/2}dq \\ &= e^{\mu+\sigma^2/2}\int_{-\infty}^{\Phi^{-1}(p)} \frac{1}{\sqrt{2\pi}} e^{-(q-\sigma)^2/2}dq \\ &= \Phi(\Phi^{-1}(p)-\sigma)e^{\mu+\sigma^2/2}.\end{aligned}$$

We have found that if Z_1 is $\mathrm{N}(\mu,\sigma^2)$-distributed, then

$$\mathrm{ES}_p(X) = V_0\left(1 - \frac{\Phi(\Phi^{-1}(p)-\sigma)e^{\mu+\sigma^2/2}}{pR_0}\right).$$

Similarly, if Z_1 is distributed as $\mu + \sigma Y$, where Y has a standard Student's t distribution with ν degrees of freedom, then $F_{Z_1}^{-1}(u) = \mu + \sigma t_\nu^{-1}(u)$ and

$$\mathrm{ES}_p(X) = V_0\left(1 - \frac{1}{pR_0}\int_0^p e^{\mu+\sigma t_\nu^{-1}(u)}du\right).$$

The integral expression can be evaluated by numerical integration.

6.4 Risk Measures Based on Utility Functions

Consider a concave and strictly increasing function u, that is, a utility function. Suppose that we consider a portfolio with value X at time 1 acceptable if it satisfies $\mathrm{E}[u(X)] \geq u(C)$ for a predetermined number C, i.e., if its certainty equivalent is at least C. Let

$$\rho_u(X) = \min\{m : \mathrm{E}[u(mR_0 + X)] \geq u(C)\}, \tag{6.17}$$

and note that $\rho_u(X)$ is the smallest amount of money that needs to be added and invested in a risk-free asset to make the corresponding position acceptable. In fact, $\rho_u(X)$ is the unique number m satisfying $\mathrm{E}[u(mR_0 + X)] = u(C)$. Let us prove this claim. Since u is strictly increasing, the function $m \mapsto \mathrm{E}[u(mR_0 + X)]$ is also strictly increasing, so there is at most one such number m. Since $m \mapsto u(mR_0+x)$ is concave, then $m \mapsto \mathrm{E}[u(mR_0+X)]$ is also concave and, therefore, also continuous. Therefore, there is at least one such number m.

Proposition 6.7. *The risk measure ρ_u in (6.17) is a convex measure of risk.*

Before proving the proposition we remark that ρ_u is in general not a coherent measure of risk. If $C = 0$, then the normalization property holds, but ρ_u will typically not be positively homogeneous.

Proof. We have

$$\begin{aligned}\rho_u(X + yR_0) &= \min\{m : \mathrm{E}[u((m + y)R_0 + X)] \geq u(C)\}\\ &= \min\{k : \mathrm{E}[u(kR_0 + X)] \geq u(C)\} - y\\ &= \rho_u(X) - y,\end{aligned}$$

which shows that ρ_u is translation invariant. Since u is increasing, $X_2 \leq X_1$ implies that $\mathrm{E}[u(mR_0 + X_2)] \leq \mathrm{E}[u(mR_0 + X_1)]$, and therefore

$$\begin{aligned}\rho_u(X_2) &= \min\{m : \mathrm{E}[u(mR_0 + X_2)] \geq u(C)\}\\ &\geq \min\{m : \mathrm{E}[u(mR_0 + X_1)] \geq u(C)\}\\ &= \rho_u(X_1),\end{aligned}$$

which proves the monotonicity of ρ_u. By the definition of ρ_u, it holds that

$$\rho_u(\lambda X_1 + (1 - \lambda)X_2) = \min\{m : \mathrm{E}[u(mR_0 + \lambda X_1 + (1 - \lambda)X_2)] \geq u(C)]\}.$$

Therefore, the convexity of ρ_u follows if we show that $m_0 = \lambda\rho_u(X_1) + (1 - \lambda)\rho_u(X_2)$ satisfies $\mathrm{E}[u(m_0R_0 + \lambda X_1 + (1 - \lambda)X_2)] \geq u(C)$. Indeed,

$$\begin{aligned}&\mathrm{E}[u([\lambda\rho_u(X_1) + (1 - \lambda)\rho_u(X_2)]R_0 + \lambda X_1 + (1 - \lambda)X_2)]\\ &\quad\geq \lambda\,\mathrm{E}[u(\rho_u(X_1)R_0 + X_1)] + (1 - \lambda)\,\mathrm{E}[u(\rho_u(X_2)R_0 + X_2)]\\ &\quad= \lambda u(C) + (1 - \lambda)u(C)\\ &\quad= u(C),\end{aligned}$$

where the first inequality holds because u is concave and where the second to last equality holds because $\mathrm{E}[u(\rho_u(X_k)R_0 + X_k)] = u(C)$ by definition of ρ_u. The proof is complete. □

6.5 Spectral Risk Measures

Consider a random variable X representing the value at time 1 of a portfolio. Let R_0 be the return of a zero-coupon bond maturing at time 1, and let F_{X/R_0} be the distribution function of X/R_0, i.e., the discounted future portfolio value. A natural set of risk measures consists of risk measures that can be written as

-1 times a weighted average of the quantile values $F^{-1}_{X/R_0}(p)$. We have seen that $\mathrm{VaR}_p(X) = -F^{-1}_{X/R_0}(p)$ for those x where $F_{X/R_0}(x)$ is neither flat nor has a jump and that

$$\mathrm{ES}_p(X) = -\frac{1}{p}\int_0^p F^{-1}_{X/R_0}(u)du.$$

In particular, ES_p puts equal weight on all the quantiles $F^{-1}_{X/R_0}(u)$ for $u < p$. It is not at all evident that this is the most natural choice. Consider a nonnegative function ϕ on $(0, 1)$ that is decreasing and integrates to 1, and define

$$\rho_\phi(X) = -\int_0^1 \phi(u) F^{-1}_{X/R_0}(u)du. \tag{6.18}$$

A risk measure ρ_ϕ with this representation is called a spectral risk measure, and the function ϕ is called the risk aversion function. A tractable property of spectral risk measures is that, like risk measures based on utility functions, all quantile values of the probability distribution of the considered portfolio value can be taken into account—not just those corresponding to the left tail. We see that ES_p is a spectral risk measure with risk aversion function $p^{-1}I_{(0,p)}$. This risk aversion function says that the worst fractions p of quantile values are weighted equally as they enter only through their mean value. In particular, extreme losses are not considered worse (receive higher weights) than less extreme losses. In general, the risk aversion function lets you specify your attitude toward risk. In spirit, it is similar to a utility function. The difference is that the utility function relates how much you value x units of cash over y units of cash, whereas the risk aversion function relates how highly you penalize the quantile at level p over the quantile at level q. Two examples of risk aversion functions are the polynomial and exponential risk aversion functions given by

$$\phi_{\mathrm{pol},\beta}(p) = \frac{1}{\beta}(1-p)^{\beta-1}, \quad \beta \geq 1,$$

$$\phi_{\exp,\gamma}(p) = \frac{\gamma \exp\{-\gamma p\}}{1-\exp\{-\gamma\}}, \quad \gamma > 0.$$

Note that both functions are decreasing and integrate to 1. For the most part we will in the sequel assume that the risk aversion function ϕ is differentiable. This assumption is made purely for convenience. The results presented below hold also without this assumption.

We begin with two useful representations of a spectral risk measure. The first one shows, using integration by parts, that ρ_ϕ can be viewed as a weighted average of ES. The second representation is similar to representation (6.15) for ES but requires the more general convex optimization from Sect. 2.2.

Proposition 6.8. *If ϕ is differentiable, then ρ_ϕ in (6.18) satisfies*

$$\rho_\phi(X) = -\int_0^1 \frac{d\phi}{du}(u)u\,\mathrm{ES}_u(X)du - \phi(1)\,\mathrm{E}[X/R_0], , \tag{6.19}$$

$$\rho_\phi(X) = \min_f \int_0^1 \frac{d\phi}{du}(u)\{uf(u) - \mathrm{E}[(f(u) - X/R_0)_+]du - \phi(1)\,\mathrm{E}[X/R_0], \tag{6.20}$$

where the minimum is taken over all functions f.

Proof. First observe that

$$\int_0^1 F^{-1}_{X/R_0}(u)du = \mathrm{E}[X/R_0].$$

This follows, for instance, from (6.14) with $p = 1$. Then, upon changing the order of integration in the third equality below, we find that

$$\begin{aligned}
\rho_\phi(X) &= -\int_0^1 \phi(v)F^{-1}_{X/R_0}(v)dv \\
&= \int_0^1 \left[\int_v^1 \frac{d\phi}{du}(u)du - \phi(1)\right] F^{-1}_{X/R_0}(v)dv \\
&= \int_0^1 \frac{d\phi}{du}(u)\left[\int_0^u F^{-1}_{X/R_0}(v)dv\right]du - \phi(1)\int_0^1 F^{-1}_{X/R_0}(v)dv \\
&= -\int_0^1 \frac{d\phi}{du}(u)u\left[\left(-\frac{1}{u}\right)\int_0^u F^{-1}_{X/R_0}(v)dv\right]du - \phi(1)\,\mathrm{E}[X/R_0] \\
&= -\int_0^1 \frac{d\phi}{du}(u)u\,\mathrm{ES}_u(X)du - \phi(1)\,\mathrm{E}[X/R_0].
\end{aligned}$$

This proves (6.19). Informally, the second representation (6.20) follows from the representation of $\mathrm{ES}_u(X)$ in (6.15). Write

$$\mathrm{ES}_u(X) = \min_{f(u)} -f(u) + \frac{1}{u}\,\mathrm{E}[(f(u) - X/R_0)_+],$$

insert this expression into (6.19), and finally move the coordinatewise minimum inside the integral out of the integral to get (6.20). Now we consider a more formal argument, in the context of Sect. 2.2. Let

$$F(f) = \int_0^1 \frac{d\phi}{du}(u)\{uf(u) - \mathrm{E}[(f(u) - X/R_0)_+]du - \phi(1)\,\mathrm{E}[X/R_0].$$

Since there are no constraints on f, here we have

$$H(f,g) = \int_0^1 \frac{d\phi}{du}(u)\{u - F_{X/R_0}(f(u))\}(g(u) - f(u))du.$$

If $H(f,g) = 0$ for each g, then f must satisfy $F_{X/R_0}(f(u)) = u$ for each u. However, for such a function f it follows, as in the proof of (6.15), that

$$-f(u) + \frac{1}{u}E[(f(u) - X/R_0)_+] = \mathrm{ES}_u(X),$$

and therefore it follows from (6.19) that the minimum of $F(f)$ is given by

$$-\int_0^1 \frac{d\phi}{du}(u)u\,\mathrm{ES}_u(X)du - \phi(1)\,\mathrm{E}[X/R_0] = \rho_\phi(X).$$

The proof is complete. □

From representation (6.19) we observe that many properties of spectral risk measures follow from properties of ES. In particular, spectral risk measures are coherent.

Proposition 6.9. *The spectral risk measure ρ_ϕ in* (6.18) *is a coherent measure of risk.*

Proof. Since ϕ is nonnegative and integrates to 1, the properties of the quantile function imply that ρ_ϕ is translation invariant, monotone, and positively homogeneous. To prove subadditivity, we make the additional assumption that the risk aversion function ϕ is differentiable. Then the subadditivity of ρ_ϕ follows from the subadditivity of ES. Indeed, for two future portfolio values X_1 and X_2 we have

$$\begin{aligned}
\rho_\phi(X_1 + X_2) &= -\int_0^1 \frac{d\phi}{du}(u)u\,\mathrm{ES}_u(X_1 + X_2)du - \phi(1)\,\mathrm{E}[(X_1 + X_2)/R_0] \\
&\le -\int_0^1 \frac{d\phi}{du}(u)u\Big(\mathrm{ES}_u(X_1) + \mathrm{ES}_u(X_2)\Big)du - \phi(1)\,\mathrm{E}[(X_1 + X_2)/R_0] \\
&= \rho_\phi(X_1) + \rho_\phi(X_2).
\end{aligned}$$

□

6.6 Notes and Comments

An extensive account of VaR for financial risk management is given in the book [24] by Philippe Jorion. The concept of coherent measures of risk was proposed by Philippe Artzner, Freddy Delbaen, Jean-Marc Eber and David Heath [4]. For an extensive account of convex and coherent measures of risk see the book [17] by Hans Föllmer and Alexander Schied. The coherence of ES was proved by Carlo Acerbi and Dirk Tasche in [2]. An introduction to and properties of spectral risk measures can be found in Acerbi's work [1]. Portfolio optimization with ES constraints was considered by Tyrrell Rockafellar and Stan Uryasev in [38] and extended to so-called generalized deviations, which are closely related to spectral risk measures, in works by Rockafellar, Uryasev, and Michael Zabarankin [39–41].

6.7 Exercises

In the exercises below, it is assumed, wherever applicable, that you can take positions corresponding to fractions of assets.

Exercise 6.1 (Convexity and subadditivity). Show that a positively homogeneous risk measure is convex if and only if it is subadditive.

Exercise 6.2 (Stop-loss reinsurance). Suppose that the total claim amount S in 1 year for an insurance company has a standard exponential distribution. The insurance company can buy so-called stop-loss reinsurance so that a claim amount exceeding $F_S^{-1}(0.95)$ is paid by the reinsurer. In this case, the insurance company has to pay $L = \min(S, F_S^{-1}(0.95)) + p$, where p is the premium paid for the stop-loss reinsurance. Determine the premium p for which $F_S^{-1}(0.99) = F_L^{-1}(0.99)$.

Exercise 6.3 (Quantile bound). Let Z denote the daily log return of an asset. Empirical studies suggest that Z has zero mean, standard deviation 0.01, and a symmetric density function. Someone claims that $F_Z^{-1}(0.99) = 0.1$. Use Chebyshev's inequality $\mathrm{P}(|Z - \mathrm{E}[Z]| > x) \leq x^{-2}\,\mathrm{Var}(Z)$, for $x > 0$, to show that this claim is false.

Exercise 6.4 (Tail conditional median). The tail conditional median $\mathrm{TCM}_p(X) = \mathrm{median}[L \mid L \geq \mathrm{VaR}_p(X)]$, where $L = -X/R_0$, has been proposed as a more robust alternative to $\mathrm{ES}_p(X)$ since $\mathrm{TCM}_p(X)$ is not as sensitive as $\mathrm{ES}_p(X)$ to the behavior of the left tail of the distribution of X.

Let Y have a standard Student's t distribution with ν degrees of freedom, and set $X = e^{0.01Y} - 1$. Compute and plot the graphs of $\mathrm{ES}_{0.01}(X)$ and $\mathrm{TCM}_{0.01}(X)$ as functions of $\nu \in [1, 15]$.

Exercise 6.5 (Production planning). Consider a company that has the option to start production of a volume $t \geq 0$ of a certain good during the next year. The company has capital of \$10,000 to use for the production. Any capital not spent on production is deposited in a bank account that does not pay interest. The cost for producing a volume $t > 0$ of the good is t thousand dollars plus a startup cost of \$5,000. The income from selling a volume t of the good is $5t$ thousand dollars. The unknown demand for the good (the maximum volume the company can sell) is modeled as a random variable with distribution function $1 - x^{-2}$, $x \geq 1$.

(a) How much should the company produce to maximize $E[V_1(t)]$, where $V_1(t)$ is the income from sales plus money in the bank account at the end of next year when producing volume t of the good?
(b) Compute $\mathrm{VaR}_p(V_1(t) - 10{,}000)$ with $V_1(t)$ as in (a) and where t is the maximizer of $E[V_1(t)]$.

Exercise 6.6 (Risky bonds). Consider a market with an asset with a risk-free 1-year return of $R_0 = 1.05$. There are also two defaultable bonds on the market whose issuers can be assumed to default independently of each other. Both bonds have maturity in 1 year and a face value of \$100,000, which is paid in the case of no default before the end of the year. For each bond a default event makes the bond worthless. Both bonds have the same price of $100{,}000(1 - q)/R_0$ dollars today, where $q = 0.025$ can be interpreted as the market's implied default probability. You believe that the market is overestimating the default probability, which you believe is $p = 0.024$. You have $V_0 = \$1$ million to invest in the risky bonds and in the risk-free asset.

(a) Determine the portfolio that maximizes your expected return given that the standard deviation of your portfolio does not exceed \$25,000. You are not allowed to take short positions in the risky bonds or in the risk-free asset.
(b) Determine the expected value and the standard deviation of the value at the end of the year of the optimal portfolio in (a).
(c) Compute $\mathrm{VaR}_{0.05}(V_1 - V_0 R_0)$ and $\mathrm{ES}_{0.05}(V_1 - V_0 R_0)$, where V_1 is the value of the optimal portfolio in (a) at the end of the year.
(d) Shortly after you buy the portfolio, a financial crisis breaks out and you realize that one of the issuers is in serious financial distress. You update the default probability to 0.91 for one of the bonds. The other bond is unaffected by the crisis, and its default probability remains 0.024. You can assume that the default events are independent. Compute $\mathrm{VaR}_{0.05}(V_1 - V_0 R_0)$ and $\mathrm{ES}_{0.05}(V_1 - V_0 R_0)$, where V_1 is the value of the optimal portfolio in (a) at the end of the year.

Exercise 6.7 (Leverage and margin calls). Consider the portfolio in Exercise 3.3(c).

(a) Compute $\mathrm{VaR}_p(V_2)$ for $p \leq 0.05$, where V_2 is the value in 2 months of the portfolio in Exercise 3.3 (c) that maximizes the expected payoff in 2 months.
(b) Compute $\mathrm{ES}_p(V_2)$ for $p \leq 0.05$, where V_2 is as in (a).

Exercise 6.8 (Risk and diversification). Consider the setup in Example 6.10 with 100 identical bonds whose default events are independent. Consider an investor with initial capital of $V_0 = \$1$ million who invests this capital in long positions of equal size in $n \leq 100$ of the bonds. The value of the bond portfolio at maturity of the bonds is denoted by $V_1(n)$.

(a) Plot $\mathrm{VaR}_{0.05}(V_1(n) - V_0)$ as a function of n, where n ranges from 1 to 100.
(b) Plot $\mathrm{ES}_{0.05}(V_1(n) - V_0)$ as a function of n, where n ranges from 1 to 100.

Project 6 (Collar options). A private investor owns a large quantity of shares of a single stock and is worried about the position being too risky in the near future. A bank offers the investor the opportunity to implement a collar option as protection against falling share prices. The collar option considered here is a long position in a European put option on the future share price with a strike price below the current share price and a short position of the same size in a European call option with strike price above the current share price with the same time to maturity as for the put option.

Suppose that the investor holds 1,000 shares and the current share price is $100. Suppose further that the strike prices of the put and call options are $95 and $105, respectively, and both options expire in 2 months. Suppose that the stock pays no dividends within the next 2 months, that all interest rates are zero, and that the put and call prices correspond to implied volatilities of 0.25 and 0.2, respectively, per year if the Black–Scholes formulas for European put and call options are used.

Suppose that the log return of the share price from today until half a month from today is $0.04X$, where X has a standard Student's t distribution with 4 degrees of freedom, and that the implied volatilities in half a month from today are the same as today.

(a) The investor decides to take a collar option position corresponding to 1,000 puts and calls. The investor's collar option position is financed by a zero-interest-rate loan if the initial value is positive. If the value is negative, then the investor receives cash that is deposited in an account that pays no interest. Express V_1, the value in half a month from today of the shares and the collar option position minus the current value of the collar option position, as a function of the log return $0.04X$.

(b) Consider the same situation as in (a) and compute $\mathrm{VaR}_{0.05}(V_1 - V_0)$, where V_0 is the current value of the shares. Compare the result to the corresponding result in a situation where the investor decides not to take a collar option position (only shares).

(c) The investor decides to take a collar option position corresponding to $h \in [0, 1{,}000]$ puts and calls. Vary h and study the effect on the density function of V_1, where V_1 is the value in half a month from today of the shares and the collar option position minus the current value of the collar option position.