Chapter 2
Risk Measures

2.1 The Notion of a Risk Measure

In ordinary language, risk is simply understood as the possibility that "unfavorable events" occur. Deviations toward the positive are as a rule ignored. When one tries to capture "risk" quantitatively it turns out that risk is very much a many-sided phenomenon.

One way to describe risk mathematically consists in identifying risk in general with fluctuations (for example monetary fluctuations). In that way both "favorable" and "unfavorable" variations are taken into account. Such an approach is taken, for example, when one chooses as a measure of risk the standard deviation (see below).

Another focus would be to identify financial risks with an amount of money, which gives an indication of how much one can lose in the event of the risk occurring. This will be the approach that we will mainly follow. For this, various measures are suitable according to the situation. Very popular are measures whose results can be viewed operationally as the amount of capital that the company must put by, according to its level of risk aversion, in order to go about its business. The result from such a measure is referred to as the *risk capital*.

Let $(\Omega, \mathscr{A}, \mathbf{P})$ be a probability space with a σ-algebra $\mathscr{A}$ and probability measure $\mathbf{P}$. We denote by $\mathscr{M}_{\mathscr{B}}(\Omega, \mathbb{R}^k)$ the space of $\mathbb{R}^k$-valued random variables

$$X : \Omega \to \mathbb{R}^k, \quad \omega \mapsto X(\omega),$$

that is the maps measurable with respect to $\mathscr{A}$ and the Borel σ-algebra. If we wish to emphasize the σ-algebra $\mathscr{A}$, we will also say *$\mathscr{A}$-measurable maps* or the *maps measurable with respect to $\mathscr{A}$*.

Definition 2.1 A *risk measure* is a map

$$\rho : \mathscr{M}(\Omega, \mathbb{R}) \to \mathbb{R}, \quad X \mapsto \rho(X),$$

where $\mathscr{M}(\Omega, \mathbb{R}) \subseteq \mathscr{M}_{\mathscr{B}}(\Omega, \mathbb{R})$ is suitable vector subspace (depending on ρ).

M. Kriele, J. Wolf, *Value-Oriented Risk Management of Insurance Companies*, EAA Series, DOI 10.1007/978-1-4471-6305-3_2, © Springer-Verlag London 2014

Remark 2.1 The restriction to a subspace is necessary since the interesting risk measures are often not defined on all of $\mathscr{M}_{\mathscr{B}}(\Omega, \mathbb{R}^k)$. If, in the following, we use the notation $\mathscr{M}(\Omega, \mathbb{R}^k)$, it is always intended in the context of an obviously suitable subspace of $\mathscr{M}_{\mathscr{B}}(\Omega, \mathbb{R}^k)$.

2.2 Examples of Risk Measures

Let Y be a random variable that describes an uncertain financial outcome. Then $X = -Y$ expresses the possible loss. Many risk measures contain a parameter $\alpha \in]0,1[$, through which the (intuitive) safety levels described by this measure are fixed. Here we shall call this parameter a *confidence level* and reserve the notion of *safety level* for its intuitive meaning. Safety level is made mathematically concrete by giving a risk measure, a confidence level, and the time horizon, to which the profit and loss amounts refer. The terminology in the literature is however fairly confused, so that the actual meaning is only clear in context.

2.2.1 Measures Based on Moments

2.2.1.1 Measures Based on the Standard Deviation

A mathematically very simple measure of risk is the standard deviation,

$$\sigma(X) = \sqrt{\mathrm{E}\big((X - \mathrm{E}(X))^2\big)} = \sqrt{\mathrm{E}\big((Y - \mathrm{E}(Y))^2\big)} = \sqrt{\mathrm{var}(X)} = \sqrt{\mathrm{var}(Y)}.$$

It expresses how far on average the results differ from the expected value, where the "difference measure" is simply borrowed from Euclidean geometry. As a measure of risk, standard deviation is also used in the form

$$\rho(X) = a\,\mathrm{E}(X) + b\sigma(X) \tag{2.1}$$

where $a, b > 0$ are given parameters. A traditional area of application of this measure is in fixing premiums. A related principle applied in fixing premiums is the *variance principle* with the measure of risk

$$\rho(X) = a\,\mathrm{E}(X) + b\sigma^2(X). \tag{2.2}$$

In the variance principle it is to be noted that the variance does not represent, as the expected value does, an amount of money, but rather a quadratic amount of money, so that the sum of $a\,\mathrm{E}(X)$ and $b\sigma^2(X)$ is difficult to interpret.

The risk measure (2.1) has the unpleasant property that positive variations in the standard deviation have the same effect as negative variations. It is thus insensitive to whether an event is "favorable" or "unfavorable". To avoid this problem one can consider only the losses that exceed the expected value by using the one-sided standard deviation $\sigma_+ = \sqrt{\mathrm{E}(\max(0, X - \mathrm{E}(X)))^2}$.

2.2.1.2 Risk Measures Based on Higher Moments

Risk measures that are based only on the mean and standard deviation do not take into account that loss distributions are in general very asymmetric. Examples of this are the claim amount distributions in property insurance and surplus sharing in life insurance with guaranteed interest. This asymmetry can be taken into account by including higher moments in the risk measure.

2.2.1.3 Shortfall Measures

The danger of exceeding a given loss threshold a is measured by so-called shortfall measures. The higher and lower partial moments weight the discrepancy for this with a power function.

For loss amounts one considers the *upper partial moments*:

$$\mathrm{UPM}_{(h,a)}(X) = \begin{cases} \mathrm{E}(\max(0, X-a)^h) & \text{for } h > 0 \\ \mathbf{P}(X \geq a) & \text{for } h = 0. \end{cases}$$

Special cases are the probability of exceeding the critical boundary a $(h = 0)$, the mean excess $(h = 1)$ and the semi-variance $(h = 2)$.

For profit amounts there are analogously the *lower partial moments*:

$$\mathrm{LPM}_{(h,a)}(Y) = \begin{cases} \mathrm{E}(\max(0, a-Y)^h) & \text{for } h > 0 \\ \mathbf{P}(Y \leq a) & \text{for } h = 0. \end{cases}$$

2.2.1.4 General Problems with Moment-Based Measures

The biggest problem with moment-based measures is the fact that they can be interpreted financially only with difficulty. Most convincingly the standard deviation can be seen as the "average distance to the mean". However, the Euclidean metric, though a good measure of distance, is no natural measure of financial risks.

For many of the distributions used in the insurance industry higher moments do not exist. In modeling operational risks using the GPD (generalized Pareto distribution) with values of parameters that may occur in practice not even the mean is defined.[1] Moment-based measures cannot be applied to such distributions.

[1] In such a case the company cannot last indefinitely, if the management of operational risks is not significantly improved.

2.2.2 Value at Risk

In contrast, the value at risk is a direct and simple mathematical financial measure. It describes the amount that, with a given probability α, one may at most “lose”.

Definition 2.2 The *value at risk* (or *VaR* for short) $\mathrm{VaR}_\alpha(X)$ is given by the formula

$$\mathrm{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\},$$

where F_X is the distribution function of X.

The value at risk, $\mathrm{VaR}_\alpha(X)$, is the *minimal* loss, that occurs in $100(1-\alpha)$ % of the worst scenarios for the portfolio (see Fig. 2.1).

In other words, if a company does not wish to consume, with the probability α, all its equity capital in a time period, then that equity capital must amount to *at least* $\mathrm{VaR}_\alpha(X)$, where X denotes the loss in this time interval. This measure is therefore suitable for a shareholder who is only liable for the money he has invested. For internal risk management, where one is also interested in higher risks beyond the quantile $\mathrm{VaR}_\alpha(X)$, this measure is not always suitable.

Remark 2.2 In exceptional cases $\mathrm{VaR}_\alpha(X)$ can be negative for large α. This value then would correspond to a profit not a loss.

In the language of statistics, value at risk represents the lower α-quantile of the distribution of X. In the special case that F_X is invertible, we have $\mathrm{VaR}_\alpha(X) = F_X^{-1}(\alpha)$.

Lemma 2.1 *For all* $\alpha \in]0,1[$ *we have* $F_X(\mathrm{VaR}_\alpha(X)) = \alpha$.

Proof This follows directly from the right continuity of the distribution function. □

The following two lemmas make clear that value at risk can be considered as “pseudo-inverse” of the distribution function of X.

Lemma 2.2 *If* F_X *is the distribution function of* X, *then* $\mathrm{VaR}_{F(X)}(X) = X$ *a.e.*

Proof Because of the monotony of F_X there holds

$$Y := \mathrm{VaR}_{F_X \circ X}(X) = \inf\{x \in \mathbb{R} : F_X(x) \geq F_X \circ X\} \leq X \quad \text{a.e.}$$

From Lemma 2.1 follows in addition that $F_X(Y(\omega)) = F_X(X(\omega))$ for all $\omega \in \Omega$. Hence F_X is constant on each interval $[Y(\omega_0), X(\omega_0)[$ where $\omega_0 \in \{\omega : Y(\omega) < X(\omega)\}$. As a result, $\mathbf{P}(Y < X) = 0$. □

Lemma 2.3 *Let* U *be a random variable with* $\mathbf{P}(U \leq u) = u$ *for all* $u \in]0,1[$. *Then the random variable* $\mathrm{VaR}_{U(\cdot)}(X)$ *has the same distribution function as* X.

Proof Let $\omega \in \Omega$ with $U(\omega) \leq F_X(x)$. Then obviously

$$y_0 = \inf\big\{y\colon U(\omega) \leq F_X(y)\big\} \leq x,$$

since x itself satisfies the condition for y. Conversely it follows from the right continuity of F_X that the equation $U(\omega) \leq F_X(y_0)$ is also satisfied for the infimum y_0. Thus we have shown

$$\big\{\omega \in \Omega\colon U \leq F_X(x)\big\} = \big\{\omega \in \Omega\colon \inf\big\{y\colon U \leq F_X(y)\big\} \leq x\big\},$$

and it follows that

$$\begin{aligned}\mathbf{P}\big(\mathrm{VaR}_{U(\cdot)}(X) \leq x\big) &= \mathbf{P}\big(\inf\big\{y : F_X(y) \geq U\big\} \leq x\big)\\ &= \mathbf{P}\big(U \leq F_X(x)\big)\\ &= F_X(x) = \mathbf{P}(X \leq x).\end{aligned}$$

□

Lemma 2.4 *Let* $\mathscr{M}(\Omega, \mathbb{R})$ *and* $\alpha \in]0,1[$. *Then*

$$\mathbf{P}\big(X < \mathrm{VaR}_\alpha(X)\big) \leq \alpha \leq \mathbf{P}\big(X \leq \mathrm{VaR}_\alpha(X)\big).$$

If it is also true that $\mathbf{P}(X = \mathrm{VaR}_\alpha(X)) = 0$, *then, in particular, it follows that* $\alpha = \mathbf{P}(X \leq \mathrm{VaR}_\alpha(X))$.

Proof Let U be a random variable with $\mathbf{P}(U \leq u) = u$ for all $u \in]0,1[$. Since the value at risk grows monotonely with the confidence level

$$\begin{aligned}\big\{\omega\colon \mathrm{VaR}_{U(\omega)}(X) < \mathrm{VaR}_\alpha(X)\big\} &\subseteq \big\{\omega : U(\omega) < \alpha\big\}\\ &\subseteq \big\{\omega\colon \mathrm{VaR}_{U(\omega)}(X) \leq \mathrm{VaR}_\alpha(X)\big\}.\end{aligned}$$

From Lemma 2.3 now follows

$$\begin{aligned}\mathbf{P}\big(X < \mathrm{VaR}_\alpha(X)\big) &= \mathbf{P}\big(\mathrm{VaR}_{U(\cdot)}(X) < \mathrm{VaR}_\alpha(X)\big)\\ &\leq \overbrace{\mathbf{P}\big(U(\cdot) < \alpha\big)}^{=\alpha}\\ &\leq \mathbf{P}\big(\mathrm{VaR}_{U(\cdot)}(X) \leq \mathrm{VaR}_\alpha(X)\big) = \mathbf{P}\big(X \leq \mathrm{VaR}_\alpha(X)\big).\end{aligned}$$

Under the additional assumption that $\mathbf{P}(X = \mathrm{VaR}_\alpha(X)) = 0$, the inequalities degenerate into equalities, since $\mathbf{P}(X < \mathrm{VaR}_\alpha(X)) = \mathbf{P}(X \leq \mathrm{VaR}_\alpha(X))$ holds. □

For the important class of normal random variables the value at risk can be given directly:

Proposition 2.1 *Let* $X\colon \Omega \to \mathbb{R}$ *be a normally distributed random variable with mean* m *and standard deviation* s. *If* $\Phi_{0,1}$ *is the standard normal distribution and*

$f: X(\Omega) \to \mathbb{R}$ is a strongly monotone increasing map, then

$$\mathrm{VaR}_\alpha(f \circ X) = f\big(m + s{\Phi_{0,1}}^{-1}(\alpha)\big).$$

Proof Since $F_{f\circ X}$ is strongly monotone increasing the value at risk is uniquely determined by $F_{f\circ X}(\mathrm{VaR}_\alpha(f \circ X)) = \alpha$. The assertion thus follows from

$$\begin{aligned}\mathbf{P}\big(f \circ X \le f\big(m + s{\Phi_{0,1}}^{-1}(\alpha)\big)\big) &= \mathbf{P}\big(X \le m + s{\Phi_{0,1}}^{-1}(\alpha)\big) \\ &= \mathbf{P}\left(\frac{X-m}{s} \le {\Phi_{0,1}}^{-1}(\alpha)\right) \\ &= \Phi_{0,1}\big({\Phi_{0,1}}^{-1}(\alpha)\big) = \alpha,\end{aligned}$$

where we have used that f is invertible on $X(\Omega)$ and $\omega \mapsto \frac{X-m}{s}$ has a standard normal distribution. □

Example 2.1 If X is log-normally distributed with the parameters m and s^2, then there holds $\mathrm{VaR}_\alpha(X) = \exp(m + s\Phi_{0,1}^{-1}(\alpha))$.

2.2.3 Tail Value at Risk and Expected Shortfall

The tail value at risk, in contrast to the value at risk, also weights higher losses.

Definition 2.3 The *tail value at risk* is given by the conditional expectation

$$\mathrm{TailVaR}_\alpha(X) = \mathrm{E}\big(X \mid X > \mathrm{VaR}_\alpha(X)\big).$$

It thus delivers, from the internal risk management point of view, more interesting information, namely the *expected* loss of the $100(1-\alpha)\%$ worst scenarios. It is clear that the tail value at risk for the same confidence level α is always bigger than (or in the extreme case equal to) the value at risk. See Figs. 2.1 and 2.2.

The tail value at risk has an economic interpretation. For continuous distributions X_1, X_2 it also has, as we shall see later, the important property of subadditivity

$$\mathrm{TailVaR}_\alpha(X_1 + X_2) \le \mathrm{TailVaR}_\alpha(X_1) + \mathrm{TailVaR}_\alpha(X_2),$$

which intuitively expresses that the risk in a diversified collective is less than the sum of the individual risks. This property does not hold in general for random variables X_1, X_2 with distribution functions that have discontinuities (jumps). In contrast, the closely related 'expected shortfall' shows subadditivity for all random variables (see Sect. 2.3).

Definition 2.4 The *expected shortfall* is given by the formula

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_z(X)\,\mathrm{d}z.$$

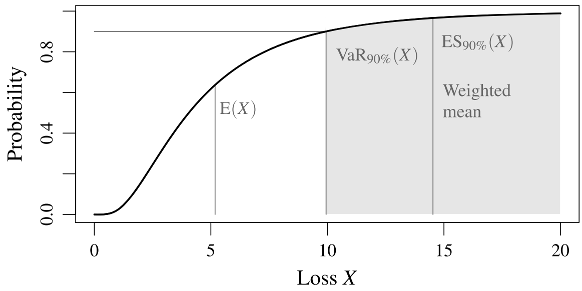

Fig. 2.1 Value at risk and tail value at risk from the perspective of the distribution function

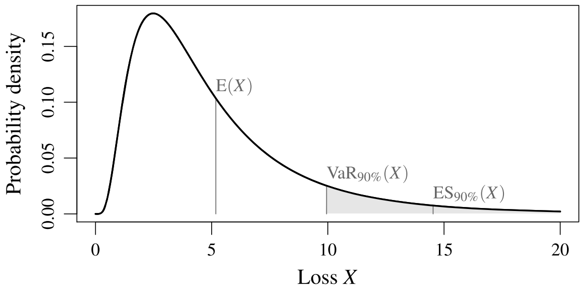

Fig. 2.2 Value at risk and tail value at risk from the perspective of the density

In the literature, the expected shortfall is sometimes called the *average value at risk*.

We will now derive an alternative formula for $\mathrm{ES}_\alpha(X)$ which shows that for continuous distribution functions $\mathrm{ES}_\alpha(X)$ coincides with $\mathrm{TailVaR}_\alpha(X)$.

Lemma 2.5 *Let $X: \Omega \to \mathbb{R}$ be a random variable and assume $x \in \mathbb{R}$. We put*

$$1_{X,x,\alpha} = 1_{\{X>x\}} + \beta_{X,\alpha}(x) 1_{\{X=x\}},$$

where

$$\beta_{X,\alpha}(x) = \begin{cases} \frac{\mathbf{P}(X \le x) - \alpha}{\mathbf{P}(X=x)} & \text{if } \mathbf{P}(X=x) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

(i) $1_{X,\mathrm{VaR}_\alpha(X),\alpha}(\omega) \in [0,1]$ *for all* $\omega \in \Omega$,
(ii) $\mathrm{E}(1_{X,\mathrm{VaR}_\alpha(X),\alpha}) = 1 - \alpha$,
(iii) $\mathrm{E}(X 1_{X,\mathrm{VaR}_\alpha(X),\alpha}) = (1-\alpha)\,\mathrm{ES}_\alpha(X)$.

Proof Some quantities used in the proof are illustrated in Fig. 2.3. (i) The assertion is obvious in the special cases $\mathbf{P}(X = \mathrm{VaR}_\alpha(X)) = 0$ and $\omega \notin \{X = \mathrm{VaR}_\alpha(X)\}$. By applying Lemma 2.4 twice, we obtain

$$\begin{aligned} 0 &\le \mathbf{P}\big(X \le \mathrm{VaR}_\alpha(X)\big) - \alpha \\ &= \mathbf{P}\big(X = \mathrm{VaR}_\alpha(X)\big) + \mathbf{P}\big(X < \mathrm{VaR}_\alpha(X)\big) - \alpha \end{aligned}$$

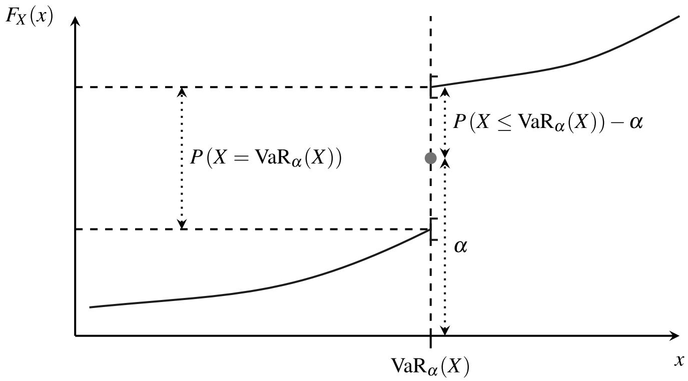

Fig. 2.3 For the proof of Lemma 2.5

$$\le \mathbf{P}\big(X = \mathrm{VaR}_\alpha(X)\big).$$

If $\mathbf{P}(X = \mathrm{VaR}_\alpha(X)) > 0$, it follows that for $\omega \in \{X = \mathrm{VaR}_\alpha(X)\}$

$$1_{X,\mathrm{VaR}_\alpha(X),\alpha}(\omega) = \frac{\mathbf{P}(X \le \mathrm{VaR}_\alpha(X)) - \alpha}{\mathbf{P}(X = \mathrm{VaR}_\alpha(X))} \in [0, 1].$$

(ii) We first consider the case $\mathbf{P}(X = \mathrm{VaR}_\alpha(X)) = 0$. Then Lemma 2.4 implies

$$\begin{aligned} \mathrm{E}(1_{X,\mathrm{VaR}_\alpha(X),\alpha}) &= \mathrm{E}(1_{\{X > \mathrm{VaR}_\alpha(X)\}}) \\ &= \mathbf{P}\big(\{X > \mathrm{VaR}_\alpha(X)\}\big) \\ &= 1 - \mathbf{P}\big(\{X \le \mathrm{VaR}_\alpha(X)\}\big) = 1 - \alpha. \end{aligned}$$

For the case $\mathbf{P}(X = \mathrm{VaR}_\alpha(X)) > 0$ we obtain

$$\begin{aligned} \mathrm{E}(1_{X,\mathrm{VaR}_\alpha(X),\alpha}) &= \mathrm{E}\left(1_{\{X > \mathrm{VaR}_\alpha(X)\}} + \frac{\mathbf{P}(X \le \mathrm{VaR}_\alpha(X)) - \alpha}{\mathbf{P}(X = \mathrm{VaR}_\alpha(X))} 1_{\{X = \mathrm{VaR}_\alpha(X)\}}\right) \\ &= \mathbf{P}\big(X > \mathrm{VaR}_\alpha(X)\big) + \frac{\mathbf{P}(X \le \mathrm{VaR}_\alpha(X)) - \alpha}{\mathbf{P}(X = \mathrm{VaR}_\alpha(X))} \mathbf{P}\big(X = \mathrm{VaR}_\alpha(X)\big) \\ &= \mathbf{P}\big(X > \mathrm{VaR}_\alpha(X)\big) + \mathbf{P}\big(X \le \mathrm{VaR}_\alpha(X)\big) - \alpha = 1 - \alpha. \end{aligned}$$

(iii) Let U be a random variable with $\mathbf{P}(U \le u) = u$ for all $u \in]0, 1[$. Since $u \mapsto \mathrm{VaR}_u(X)$ is monotone increasing, we have

$$\{U \ge \alpha\} \subseteq \left\{\mathrm{VaR}_{U(\cdot)}(X) \ge \mathrm{VaR}_\alpha(X)\right\}.$$

If $U(\omega) < \alpha$ and $\mathrm{VaR}_{U(\omega)}(X) \ge \mathrm{VaR}_\alpha(X)$, then we must have (likewise because of monotony) $\mathrm{VaR}_{U(\omega)} = \mathrm{VaR}_\alpha(X)$. Thus we also obtain the relationship

$$\{U < \alpha\} \cap \left\{\mathrm{VaR}_{U(\cdot)}(X) \geq \mathrm{VaR}_\alpha(X)\right\} \subseteq \left\{\mathrm{VaR}_{U(\cdot)}(X) = \mathrm{VaR}_\alpha(X)\right\}.$$

Overall we have derived

$$\left\{\mathrm{VaR}_{U(\cdot)}(X) \geq \mathrm{VaR}_\alpha(X)\right\} = \{U \geq \alpha\} \cup \left(\left\{\mathrm{VaR}_{U(\cdot)}(X) \geq \mathrm{VaR}_\alpha(X)\right\} \cap \{U < \alpha\}\right)$$

where $\mathrm{VaR}_{U(\omega)}(X) = \mathrm{VaR}_\alpha(X)$ for all $\omega \in \{\mathrm{VaR}_{U(\cdot)}(X) \geq \mathrm{VaR}_\alpha(X)\} \cap \{U < \alpha\}$. From this and Lemma 2.3 follows

$$\begin{aligned}
\int_\alpha^1 \mathrm{VaR}_u(X)\,\mathrm{d}u &= \mathrm{E}\big(\mathrm{VaR}_{U(\cdot)}(X)1_{\{U\geq\alpha\}}\big)\\
&= \mathrm{E}\big(\mathrm{VaR}_{U(\cdot)}(X)(1_{\{\mathrm{VaR}_{U(\cdot)}(X)\geq \mathrm{VaR}_\alpha(X)\}}\\
&\quad - 1_{\{\mathrm{VaR}_{U(\cdot)}(X)\geq \mathrm{VaR}_\alpha(X)\}\cap\{U<\alpha\}})\big)\\
&= \mathrm{E}\big(\mathrm{VaR}_{U(\cdot)}(X)1_{\{\mathrm{VaR}_{U(\cdot)}(X)\geq \mathrm{VaR}_\alpha(X)\}}\big)\\
&\quad - \mathrm{VaR}_\alpha(X)\,\mathrm{E}(1_{\{\mathrm{VaR}_{U(\cdot)}(X)\geq \mathrm{VaR}_\alpha(X)\}\cap\{U<\alpha\}})\\
&= \mathrm{E}(X1_{\{X\geq \mathrm{VaR}_\alpha(X)\}}) - \mathrm{VaR}_\alpha(X)\,\mathrm{E}(1_{\{\mathrm{VaR}_{U(\cdot)}(X)\geq \mathrm{VaR}_\alpha(X)\}\setminus\{U\geq\alpha\}})\\
&= \mathrm{E}(X1_{\{X> \mathrm{VaR}_\alpha(X)\}}) + \mathrm{E}(X1_{\{X= \mathrm{VaR}_\alpha(X)\}})\\
&\quad - \mathrm{VaR}_\alpha(X)\,\mathrm{E}(1_{\{\mathrm{VaR}_{U(\cdot)}(X)\geq \mathrm{VaR}_\alpha(X)\}}) + \mathrm{VaR}_\alpha(X)\,\mathrm{E}(1_{\{U\geq\alpha\}})\\
&= \mathrm{E}(X1_{\{X> \mathrm{VaR}_\alpha(X)\}})\\
&\quad + \mathrm{VaR}_\alpha(X)\big(\mathbf{P}\big(X = \mathrm{VaR}_\alpha(X)\big) - \mathbf{P}\big(X \geq \mathrm{VaR}_\alpha(X)\big) + 1 - \alpha\big)\\
&= \mathrm{E}(X1_{\{X> \mathrm{VaR}_\alpha(X)\}}) + \mathrm{VaR}_\alpha(X)\big(\mathbf{P}\big(X \leq \mathrm{VaR}_\alpha(X)\big) - \alpha\big)\\
&= \mathrm{E}(X1_{X,\mathrm{VaR}_\alpha(X),\alpha}),
\end{aligned}$$

where in the last step we used that Lemma 2.4, in the special case

$$\mathbf{P}\big(X = \mathrm{VaR}_\alpha(X)\big) = 0,$$

implies the equation $\mathbf{P}(X \leq \mathrm{VaR}_\alpha(X)) - \alpha = 0$. □

Proposition 2.2 *Assume* $\alpha \in [0, 1[$. *For*

$$\lambda_\alpha = \frac{1 - \mathbf{P}(X \leq \mathrm{VaR}_\alpha(X))}{1-\alpha}$$

there holds $\lambda_\alpha \in [0, 1]$ *and*

$$\mathrm{ES}_\alpha(X) = \lambda_\alpha\,\mathrm{TailVaR}_\alpha(X) + (1-\lambda_\alpha)\,\mathrm{VaR}_\alpha(X).$$

In particular, the tail value at risk and expected shortfall coincide for continuous distributions.

Proof $\lambda_\alpha \in [0, 1]$ follows directly from Lemma 2.4. We calculate

$$\begin{aligned}(1-\alpha)\,\mathrm{ES}_\alpha(X) &= \mathrm{E}(X1_{X,\mathrm{VaR}_\alpha(X),\alpha})\\ &= \mathrm{E}(X1_{\{X>\mathrm{VaR}_\alpha(X)\}}) + \mathrm{VaR}_\alpha(X)\big(\mathbf{P}\big(X \le \mathrm{VaR}_\alpha(X)\big) - \alpha\big)\\ &= \mathbf{P}\big(X > \mathrm{VaR}_\alpha(X)\big)\,\mathrm{TailVaR}_\alpha(X)\\ &\quad + \mathrm{VaR}_\alpha(X)\big(1-\alpha-\big(1-\mathbf{P}\big(X \le \mathrm{VaR}_\alpha(X)\big)\big)\big)\\ &= (1-\alpha)\lambda_\alpha\,\mathrm{TailVaR}_\alpha(X) - (1-\alpha)\,\mathrm{VaR}_\alpha(X)(1-\lambda_\alpha).\end{aligned}$$

If X is continuous, then by Lemma 2.4, $\lambda_\alpha = 1$, so that $\mathrm{ES}_\alpha(X) = \mathrm{TailVaR}_\alpha(X)$ follows. □

In general, the expected shortfall has better mathematical properties than the tail value at risk (see Sect. 2.3). The following presentation of the expected shortfall serves as motivation in Sect. 2.4.4. Additionally it allows a simple proof of the important approximation result given in Proposition 2.4.

Definition 2.5 Let $\mathbf{P}, \mathbf{Q}$ be measures on the σ-algebra $\mathscr{A}$. Then $\mathbf{Q}$ is *absolutely continuous* with respect to $\mathbf{P}$ if $\mathbf{P}(A) = 0$ implies $\mathbf{Q}(A) = 0$ for all $A \in \mathscr{A}$. In this case we write $\mathbf{Q} \ll \mathbf{P}$.

Proposition 2.3 *Assume* $\mathscr{M}(\Omega, \mathbb{R}) \subseteq L^1(\Omega, \mathbb{R})$ *and*

$$\mathscr{W}_\alpha = \left\{\mathbf{Q}\colon \mathbf{Q} \text{ is a probability measure with } \mathbf{Q} \ll \mathbf{P} \text{ and } \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} \le \frac{1}{1-\alpha}\right\}.$$

Then for $X \in \mathscr{M}(\Omega, \mathbb{R})$ *there holds*

$$\mathrm{ES}_\alpha(X) = \sup_{\mathbf{Q}\in\mathscr{W}_\alpha} \big\{\mathrm{E}_{\mathbf{Q}}(X)\big\}.$$

Proof X is integrable with respect to $\mathbf{Q}$, because X is integrable with respect to $\mathbf{P}$, $\mathbf{Q} \ll \mathbf{P}$, and $\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}$ is bounded. The special choice of $\mathbf{Q}$ defined by

$$\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} = \frac{1}{1-\alpha}\,1_{X,\mathrm{VaR}_\alpha(X),\alpha}$$

(see Lemma 2.5) satisfies both the conditions $\mathbf{Q} \ll \mathbf{P}$ and $\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} \le (1-\alpha)^{-1}$. Since

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\,\mathrm{E}(X\,1_{X,\mathrm{VaR}_\alpha(X),\alpha}) = \mathrm{E}_{\mathbf{Q}}(X)$$

holds (Lemma 2.5(iii)), there follows

$$\mathrm{ES}_\alpha(X) \le \sup_{\mathbf{R}\in\mathscr{W}_\alpha} \big\{\mathrm{E}_{\mathbf{R}}(X)\big\}.$$

Now let $\mathbf{R}$ be another probability measure that satisfies the requirements $\mathbf{R} \ll \mathbf{P}$ and $\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}} \leq (1-\alpha)^{-1}$. We must show $\mathrm{E}_{\mathbf{R}}(X) \leq \mathrm{E}_{\mathbf{Q}}(X)$. The set

$$A = \left\{\omega\colon 1_{X,\mathrm{VaR}_\alpha(X),\alpha}(\omega) > 0\right\}$$

satisfies $\mathrm{E}_{\mathbf{Q}}(1_A) = 1$. By the construction of $1_{X,\mathrm{VaR}_\alpha(X),\alpha}$ we also have $X(\omega) \leq \inf_{\tilde{\omega}\in A} X(\tilde{\omega})$ for all $\omega \in \Omega \setminus A$. From this follows the inequality

$$\begin{aligned}\mathrm{E}_{\mathbf{R}}(X) &= \mathrm{E}_{\mathbf{P}}\left(\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}} X\, 1_A\right) + \mathrm{E}_{\mathbf{P}}\left(\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}} X\, 1_{\Omega\setminus A}\right) \\ &\leq \mathrm{E}_{\mathbf{P}}\left(\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}} X\, 1_A\right) + \inf_{\tilde{\omega}\in A} X(\tilde{\omega})\mathbf{R}(\Omega \setminus A).\end{aligned}$$

From

$$\mathrm{E}_{\mathbf{P}}\left(\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\, 1_A\right) = \mathrm{E}_{\mathbf{P}}\left(\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\right) = \mathrm{E}_{\mathbf{Q}}(1) = 1$$

follows

$$\mathrm{E}_{\mathbf{P}}\left(\left(\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} - \frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}}\right) 1_A\right) = 1 - \mathbf{R}(A) = \mathbf{R}(\Omega \setminus A).$$

Since for all $\omega \in \{X > \inf_{\tilde{\omega}\in A} X(\tilde{\omega})\} \subseteq A$ it holds that

$$\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} = \frac{1}{1-\alpha} \geq \frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}},$$

on this set we have

$$X\left(\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} - \frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}}\right) \geq \inf_{\tilde{\omega}\in A} X(\tilde{\omega})\left(\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} - \frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}}\right).$$

This inequality is fulfilled trivially on $\{X = \inf_{\tilde{\omega}\in A} X(\tilde{\omega})\}$, so that because $A \subseteq \{X \geq \inf_{\tilde{\omega}\in A} X(\tilde{\omega})\}$ it holds on A. We thus deduce

$$\begin{aligned}\mathrm{E}_{\mathbf{R}}(X) &\leq \mathrm{E}_{\mathbf{P}}\left(X\,\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}}\, 1_A\right) + \inf_{\tilde{\omega}\in A} X(\tilde{\omega})\,\mathrm{E}_{\mathbf{P}}\left(\left(\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} - \frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}}\right) 1_A\right) \\ &\leq \mathrm{E}_{\mathbf{P}}\left(X\,\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}}\, 1_A\right) + \mathrm{E}_{\mathbf{P}}\left(X\left(\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} - \frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}}\right) 1_A\right) \\ &= \mathrm{E}_{\mathbf{P}}\left(X\,\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\, 1_A\right) = \mathrm{E}_{\mathbf{Q}}(X).\end{aligned}$$

□

Proposition 2.4 *Assume Y an integrable positive function and $\{X_k\}_{k\in\mathbb{N}}$ a sequence of random variables with $|X_k| \leq Y$ almost surely, that converge pointwise almost surely to the random variable X. Then it holds that* $\mathrm{ES}_\alpha(X_n) \to \mathrm{ES}_\alpha(X)$.

Proof Let $\varepsilon > 0$ and $\mathbf{Q} \in \mathscr{W}_\alpha$ with $\mathrm{E}_{\mathbf{Q}}(X) \geq \mathrm{ES}_\alpha(X) - \varepsilon$. Since for each $\mathbf{R} \in \mathscr{W}_\alpha$ the inequality $0 \leq \frac{\mathrm{d}\mathbf{R}}{\mathrm{d}\mathbf{P}} \leq \frac{1}{1-\alpha}$ holds, the sequence $\{\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} X_k\}_{k \in \mathbb{N}}$ is dominated by the integrable random variables $\frac{1}{1-\alpha} Y$. In addition, $\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} X_k$ converges almost everywhere to $\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} X$. Lebesgue's Theorem thus implies $\mathrm{E}_{\mathbf{Q}}(X_k) \to \mathrm{E}_{\mathbf{Q}}(X)$. Since $\varepsilon > 0$ was arbitrary, according to Proposition 2.3 this implies $\liminf_{k\to\infty} \mathrm{ES}_\alpha(X_k) \geq \mathrm{ES}_\alpha(X)$.

There exists a subsequence $\{X_{k_j}\}_{j \in \mathbb{N}}$ with

$$\lim_{j\to\infty} \mathrm{ES}_\alpha(X_{k_j}) = \limsup_{k\to\infty} \mathrm{ES}_\alpha(X_k).$$

Assume $\mathbf{Q}_{k_j} \in \mathscr{W}_\alpha$ with

$$\left| \mathrm{ES}_\alpha(X_{k_j}) - \mathrm{E}_{\mathbf{Q}_{k_j}}(X_{k_j}) \right| \leq \frac{1}{j}.$$

Since for each j the Radon-Nikodym derivative $\frac{\mathrm{d}\mathbf{Q}_{k_j}}{\mathrm{d}\mathbf{P}}$ is measurable and we have $0 \leq \frac{\mathrm{d}\mathbf{Q}_{k_j}}{\mathrm{d}\mathbf{P}} \leq \frac{1}{1-\alpha}$, the function $f = \limsup_{j\to\infty} \frac{\mathrm{d}\mathbf{Q}_{k_j}}{\mathrm{d}\mathbf{P}}$ is also measurable with $0 \leq f \leq \frac{1}{1-\alpha}$. The measure defined by $\frac{\mathrm{d}\tilde{\mathbf{Q}}}{\mathrm{d}\mathbf{P}} = f$ is clearly in $\mathscr{W}_\alpha$, so that we have $\mathrm{ES}_\alpha(X) \geq \mathrm{E}_{\tilde{\mathbf{Q}}}(X)$. Since the X_{n_k} converge almost everywhere to X, there holds $\limsup_{j\to\infty} \frac{\mathrm{d}\mathbf{Q}_{k_j}}{\mathrm{d}\mathbf{P}} X_{k_j} = fX$. Because $|\frac{\mathrm{d}\mathbf{Q}_{n_k}}{\mathrm{d}\mathbf{P}} X_{n_k}| \leq \frac{1}{1-\alpha} Y$ we can apply Fatou's Lemma and obtain

$$\begin{aligned}
\mathrm{E}_{\tilde{\mathbf{Q}}}(X) &= \mathrm{E}_{\mathbf{P}}(fX) \\
&= \mathrm{E}_{\mathbf{P}}\left(\limsup_{j\to\infty} \frac{\mathrm{d}\mathbf{Q}_{k_j}}{\mathrm{d}\mathbf{P}} X_{k_j} \right) \\
&\geq \limsup_{j\to\infty} \left(\mathrm{E}_{\mathbf{P}}\left(\frac{\mathrm{d}\mathbf{Q}_{k_j}}{\mathrm{d}\mathbf{P}} X_{k_j} \right) \right) \\
&\geq \limsup_{j\to\infty} \left(\mathrm{ES}_\alpha(X_{k_j}) - \frac{1}{j} \right) \\
&= \limsup_{k\to\infty} \big(\mathrm{ES}_\alpha(X_k) \big).
\end{aligned}$$

Thus it also holds that $\mathrm{ES}_\alpha(X) \geq \limsup_{k\to\infty} \mathrm{ES}_\alpha(X_k)$. □

Proposition 2.4 suggests preferring the expected shortfall over the tail value at risk. This is because for large enough n it is not possible to distinguish X_n and X. Therefore the values of the corresponding risk measures should also be practically indistinguishable. This not the case for the tail value at risk, but Proposition 2.4 shows the expected shortfall does have this property which is needed for the interpretation.

Lemma 2.6 *Let $X\colon \Omega \to \mathbb{R}$ be a normally distributed random variable with mean m and standard deviation s. Assume $f\colon X(\Omega) \to \mathbb{R}$ is a strongly monotone continuous map. If $\Phi_{0,1}$ denotes the distribution function and $\varphi_{0,1} = \frac{\mathrm{d}}{\mathrm{d}x}\Phi_{0,1}$ the density of the standard normal distribution, then there holds*

$$\mathrm{ES}_\alpha(f \circ X) = \int_{\Phi_{0,1}{}^{-1}(\alpha)}^{\infty} f(m+sx)\varphi_{0,1}(x)\,\mathrm{d}x = \mathrm{TailVaR}_\alpha(f \circ X).$$

Proof From Proposition 2.1 it follows that

$$\mathrm{ES}_\alpha(f \circ X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_p(f \circ X)\,\mathrm{d}p = \frac{1}{1-\alpha}\int_\alpha^1 f\big(m + s\Phi_{0,1}{}^{-1}(p)\big)\,\mathrm{d}p.$$

By making the substitution $p = \Phi_{0,1}(x)$ we obtain $\mathrm{d}p = \varphi_{0,1}(x)\,\mathrm{d}x$ and thus

$$\mathrm{ES}_\alpha(f \circ X) = \frac{1}{1-\alpha}\int_{\Phi_{0,1}{}^{-1}(\alpha)}^{\infty} f(m+sx)\varphi_{0,1}(x)\,\mathrm{d}x.$$

The distribution function is continuous so we have $\mathrm{ES}_\alpha(f \circ X) = \mathrm{TailVaR}_\alpha(f \circ X)$. □

In the two important special cases of normally distributed and log-normally distributed random variables, the integral above can be evaluated explicitly.

Proposition 2.5 *Let $X\colon \Omega \to \mathbb{R}$ be a normally distributed random variable with mean μ and standard deviation σ. If $\Phi_{0,1}$ denotes the distribution function and $\varphi_{0,1} = \frac{\mathrm{d}}{\mathrm{d}x}\Phi_{0,1}$ the density of the standard normal distribution, then*

$$\mathrm{ES}_\alpha(X) = \mu + \sigma\frac{\varphi_{0,1}(\Phi_{0,1}{}^{-1}(\alpha))}{1-\alpha} = \mathrm{TailVaR}_\alpha(X).$$

Proof In this case we have in Lemma 2.6 that $f(x) = x$, so that the integral simplifies to

$$\mathrm{ES}_\alpha(X) = \mu + \frac{\sigma}{1-\alpha}\int_{\Phi_{0,1}{}^{-1}(\alpha)}^{\infty} x\varphi_{0,1}(x)\,\mathrm{d}x.$$

Using the fact $\varphi'_{0,1}(x) = -x\varphi_{0,1}(x)$ we obtain

$$\mathrm{ES}_\alpha(X) = \mu - \frac{\sigma}{1-\alpha}\big[\varphi_{0,1}(p)\big]_{\Phi_{0,1}{}^{-1}(\alpha)}^{\infty} = \mu + \frac{\sigma}{1-\alpha}\varphi_{0,1}\big(\Phi_{0,1}{}^{-1}(\alpha)\big).$$ □

Proposition 2.6 *Let $X\colon \Omega \to \mathbb{R}$ be a log-normally distributed random variable, i.e., $\ln X \sim N(m, s^2)$. If $\Phi_{0,1}$ denotes the distribution function and $\varphi_{0,1} = \frac{\mathrm{d}}{\mathrm{d}x}\Phi_{0,1}$ the density of the standard normal distribution, then*

$$\mathrm{ES}_\alpha(X) = \frac{\exp(m + \frac{s^2}{2})}{1-\alpha}\Phi_{0,1}\big(s - \Phi_{0,1}{}^{-1}(\alpha)\big).$$

Proof In this case we have in Lemma 2.6 that $f(x) = \exp(x)$. Thus the integral simplifies to

$$\begin{aligned}
\mathrm{ES}_\alpha(X) &= \frac{1}{1-\alpha}\frac{1}{\sqrt{2\pi}}\int_{\Phi_{0,1}{}^{-1}(\alpha)}^{\infty} \exp(m+sx)\exp\Bigl(-\frac{1}{2}x^2\Bigr)\,\mathrm{d}x \\
&= \frac{1}{1-\alpha}\frac{1}{\sqrt{2\pi}}\int_{\Phi_{0,1}{}^{-1}(\alpha)}^{\infty} \exp\Bigl(m+\frac{s^2}{2}-\frac{1}{2}(x-s)^2\Bigr)\,\mathrm{d}x \\
&= \frac{\exp(m+\frac{s^2}{2})}{1-\alpha}\frac{1}{\sqrt{2\pi}}\int_{\Phi_{0,1}{}^{-1}(\alpha)-s}^{\infty} \exp\Bigl(-\frac{1}{2}y^2\Bigr)\,\mathrm{d}y \\
&= \frac{\exp(m+\frac{s^2}{2})}{1-\alpha}\bigl(1-\Phi_{0,1}\bigl(\Phi_{0,1}{}^{-1}(\alpha)-s\bigr)\bigr) \\
&= \frac{\exp(m+\frac{s^2}{2})}{1-\alpha}\Phi_{0,1}\bigl(s-\Phi_{0,1}{}^{-1}(\alpha)\bigr).
\end{aligned}$$

In this calculation we used, in the last equality, the symmetry of the standard normal distribution. □

2.2.4 Spectral Measures

The expected shortfall can be directly generalized to take into account individual risk aversion. Instead of averaging over all $\mathrm{VaR}_z(X)$ for $z \geq \alpha$ with a uniform weight, one can employ a more general weighting function ϕ.

Definition 2.6 Let $(A, \mathscr{A}, \mu)$ be a probability space with σ-Algebra $\mathscr{A}$ and probability measure μ. Then an integrable map $\phi\colon A \to \mathbb{R}$ is called a *weight function*, if ϕ has the following properties:

(i) $\phi(\alpha) \geq 0$ for almost every $\alpha \in A$,
(ii) $\int_A \phi(\alpha)\,\mathrm{d}\mu(\alpha) = 1$.

Definition 2.7 Let $\phi \in L^1([0, 1])$ be a weight function. The risk measure

$$M_\phi(X) = \int_0^1 \mathrm{VaR}_p(X)\phi(p)\,\mathrm{d}p$$

is called the *spectral measure* of ϕ.

With a spectral measure the risk measure is also weighted by its dependence on the unlikeliness of the occurrence of a loss. The concept of a spectral measure thus permits the representation of an individual profile of risk aversion. Obviously ES_α is an example of a spectral measure. The measure VaR can be thought of as a limit value of spectral measures, since we have $\mathrm{VaR}_\alpha(X) = \int_0^1 \mathrm{VaR}_p(X)\delta_\alpha(p)\,\mathrm{d}p$, where δ_α denotes the Dirac distribution.

2.3 Choosing a Good Risk Measure

2.3.1 Risk Measures and the Intuition of Risk

An important requirement for a good risk measure is that it describe as well as possible the risk intuition of the user. A risk measure that the user may think he understands straight away does not necessarily fulfill this requirement. We would like to illustrate this point more precisely. The following set of axioms, due to Artzner et al. [1], describes properties that correspond to our intuitive notion of risk.

Definition 2.8 A risk measure ρ is called *coherent*, if it has the following properties:

Translation invariance: $\rho(X+\alpha) = \rho(X) + \alpha$ for every $X \in \mathscr{M}(\Omega, \mathbb{R})$ and any constant α.
Positive Homogeneity: $\rho(\alpha X) = \alpha\rho(X)$ for every $X \in \mathscr{M}(\Omega, \mathbb{R})$ and any positive constant α.
Monotony: $X_1 \geq X_2$ almost everywhere $\Rightarrow \rho(X_1) \geq \rho(X_2)$ for any $X_1, X_2 \in \mathscr{M}(\Omega, \mathbb{R})$.[2]
Subadditivity: $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$ for any $X_1, X_2 \in \mathscr{M}(\Omega, \mathbb{R})$.

To see how well these axioms really express our intuition of risk, we should consider what each of these four conditions says.

Translation invariance says that certain losses must be completely covered by capital but do not influence the remaining risk: A loss of which one is certain is not a risk because it is entirely predictable. From translation invariance there follows in addition that $\rho(X - \rho(X)) = 0$. The risk capital $\rho(X)$ is thus exactly the amount of money that must be retained so as to absorb the risk at the safety level defined by ρ. In this sense risk measures that satisfy translation invariance are *acceptable* [1].

Positive homogeneity is an invariance under scaling: It is inessential whether one measures risks in cents or euros. If positive homogeneity does not hold then the arbitrarily chosen unit of currency has an influence on the amount of capital, which naturally should not be true.

Monotony means that a portfolio that shows higher losses than another portfolio in all possible situations, must mean more capital is at risk. As an example consider two identical portfolios, where however one of the portfolios features a retroactively paid discount for the premiums depending on the losses.

[2] In the original article of Artzner et.al. [1] they refer to gain $Y = -X$, thus monotony is defined differently there.

Subadditivity says that there are diversification effects resulting from the combination of risky portfolios. Subadditivity is especially intuitive for an insurer since the business model of insurance rests on the effect of diversification.[3]

Remark 2.3 Sometimes both positive homogeneity and subadditivity are viewed critically.

- It is tempting to interpret homogeneity in the sense that multiplying the insured sums in a portfolio by a factor brings with it a corresponding multiplication of the risk. This is plausible for small portfolios. However for large portfolios there are additional liquidity risks as large losses incur large payments. This would contradict positive homogeneity.

  In this argument there is the implicit assumption that scaling the insured sums of all contracts by a factor λ would also scale the loss function X by the same factor. However, this is not the case as the loss distribution of the scaled portfolio be directly affected by the additional liquidity risk and thus differ from λX.
- Regarding subadditivity consider the merger of two corporations with loss distributions X_1 and X_2. The merged corporation may have a risk capital larger than $\rho(X_1) + \rho(X_2)$ as there may be internal power struggles which increase the risk.

  Again, the solution to this conundrum is that the loss distribution X of the combined corporation satisfies $X > X_1 + X_2$, because X also has to account for these power struggles.

While we have seen that these criticisms of positive homogeneity and subadditivity are rather a criticism of a naive combination of loss distributions, risk measures that are just translation invariant, monotone and convex have been proposed as an alternative to coherent risk measures: *Convex risk measures* are defined by the requirement that for each $\alpha \in [0, 1]$ and for any two loss distributions $X_1, X_2 \in \mathscr{M}(\Omega, \mathbb{R})$ the inequality

$$\rho\big(\alpha X_1 + (1-\alpha)X_2\big) \leq \alpha\rho(X_1) + (1-\alpha)\rho(X_2)$$

holds. It is clear that convexity is a weaker condition, and follows from subadditivity and positive homogeneity.

Coherent risk measures satisfy intuitive expectations in many situations. While we do not have realistic examples, there could however be areas where what is expected of a risk measure contradicts coherence. This has to be judged case-by-case. If a risk measure does not fulfill the requirements of coherence, then it has

[3]There is a subtle distinction between pooling and diversification, whereby it is argued that the insurance business depends predominantly on pooling. The distinction is based on the idea that the pooling effect can only be achieved with costs (brokers must find insurance clients) whereas diversification is in principle free (a diversified stock portfolio costs exactly the same as an undiversified one with the same market prices). In our context in which we are only concerned with risk effects this distinction is, however, secondary.

to be judged to what extent this is the result of the situation being described, and whether this property is desirable or negligible.

The following technical theorem allows the construction of new coherent risk measures on the basis of existing coherent risk measures. We shall use it later in the proof of Theorem 2.4, in which a descriptive construction of coherent measures will be given.

Theorem 2.1 *Let $(A, \mathscr{A}, \mu)$ be a probability space with σ-Algebra $\mathscr{A}$ and probability measure μ. Let $\{\rho_\alpha\}_{\alpha\in A}$ be a family of risk measures and $\mathscr{M}$ a vector space of real-valued random variables X, such that $\rho_\alpha(X)$ are μ-almost everywhere defined and μ-integrable. If all ρ_α are translation invariant, positively homogeneous, monotone resp. subadditive, then the risk measure $\rho\colon \mathscr{M} \to \mathbb{R}$, $X \mapsto \rho(X) = \int_A \rho_\alpha(X)\,\mathrm{d}\mu(\alpha)$ also has the corresponding property.*

Proof Consider $c \in \mathbb{R}$ and arbitrary random variables X and Y.

Translation invariance:

$$\begin{aligned}\rho(X+c) &= \int_A \rho_\alpha(X+c)\,\mathrm{d}\mu(\alpha) = \int_A \big(\rho_\alpha(X)+c\big)\,\mathrm{d}\mu(\alpha)\\ &= \int_A \rho_\alpha(X)\,\mathrm{d}\mu(\alpha) + c\int_A \mathrm{d}\mu(\alpha) = \rho(X)+c,\end{aligned}$$

since μ is a probability measure.

Positive homogeneity: For $c \geq 0$ we have

$$\rho(cX) = \int_A \rho_\alpha(cX)\,\mathrm{d}\mu(\alpha) = \int_A c\rho_\alpha(X)\,\mathrm{d}\mu(\alpha) = c\rho(X).$$

Monotony: Suppose $X \geq Y$ almost everywhere. Then it follows from $\rho_\alpha(X) \geq \rho_\alpha(Y)$ that

$$\rho(X) = \int_A \rho_\alpha(X)\,\mathrm{d}\mu(\alpha) \geq \int_A \rho_\alpha(Y)\,\mathrm{d}\mu(\alpha) = \rho(Y).$$

Subadditivity:

$$\rho(X+Y) = \int_A \rho_\alpha(X+Y)\,\mathrm{d}\mu(\alpha) \leq \int_A \big(\rho_\alpha(X)+\rho_\alpha(Y)\big)\,\mathrm{d}\mu(\alpha) = \rho(X)+\rho(Y).$$

□

In general the risk measure VaR_α, which perhaps does appear at first glance to be the most plausible model for risk measurement, does not satisfy the important axiom of subadditivity. The value at risk is therefore not coherent, and so does not describe our intuition of risk to a desirable extent.

Example 2.2 Let the discrete distribution X be given by

$$\begin{cases} \mathbf{P}(X = -1) = 0.96 \\ \mathbf{P}(X = 10) = 0.04. \end{cases}$$

We can interpret $-X$ as the profit distribution for an insurance contract. The premium is set as 1. There is a loss with a probability of 4 %, and the payment in the event of a loss is always equal to 11. In this simple example costs and investment return are ignored. The business is profitable since $E(-X) = 0.56$. We are interested in the risk measure $\mathrm{VaR}_{95\,\%}$. Since losses occur only with the probability $4\,\% < 1 - 95\,\%$, we have

$$\mathrm{VaR}_{95\,\%}(X) = -1.$$

With our low confidence level there is thus no positive risk.

Consider now a second distribution $Y \sim X$ that is independent of X. The total distribution $X + Y$ is then completely described by

$$\begin{cases} \mathbf{P}(X + Y = -2) = 0.96^2 = 0.9216 \\ \mathbf{P}(X + Y = 9) = 2 \times 0.96 \times 0.04 = 0.0768 \\ \mathbf{P}(X + Y = 20) = 0.04^2 = 0.0016. \end{cases}$$

Obviously

$$\mathrm{VaR}_{95\,\%}(X + Y) = 9 > 2(-1) = \mathrm{VaR}_{95\,\%}(X) + \mathrm{VaR}_{95\,\%}(Y).$$

If one were to use the value at risk as the risk measure for this example, one would have to deduce that diversification increases the risk instead of reducing it.

There are nonetheless special cases in which the value at risk is a coherent risk measure (see Theorem 2.2). First we need to do a little preparation.

Remark 2.4 We will recall some properties of Euclidean space here that we will need for the formulation and proof of the following Lemma 2.7. We consider $\mathbb{R}^n$ with a scalar product $\langle\, ,\rangle\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. The pair $(\mathbb{R}^n, \langle\, ,\rangle)$ is called *Euclidean space* and is the basis for elementary geometry. A linear map $O\colon \mathbb{R}^n \to \mathbb{R}^n$, $u \mapsto Ou$ is said to be *orthogonal* (or *an isometry*), if $\langle Ox, Oy\rangle = \langle x, y\rangle$ holds for all $x, y \in \mathbb{R}^n$. In particular, O is invertible. The *transposed map* $O^\top$ is defined by the property $\langle Ox, y\rangle = \langle x, O^\top y\rangle$ for all $x, y \in \mathbb{R}^n$. Then $O^\top O = \mathrm{id}_{\mathbb{R}^n}$, which follows from

$$\left\langle O^\top Ox, y\right\rangle = \langle Ox, Oy\rangle = \langle x, y\rangle \quad \forall x, y \in \mathbb{R}^n.$$

$O^\top$ is itself also an orthogonal map, since

$$\left\langle O^\top Ox, O^\top Oy\right\rangle = \langle x, y\rangle = \langle Ox, Oy\rangle$$

for all $x, y \in \mathbb{R}^n$, and O is invertible.

Lemma 2.7 *Let $X\colon \Omega \to \mathbb{R}^n$ be a random variable and $\phi_X\colon \mathbb{R}^n \to \mathbb{R}$, $u \mapsto E(\mathrm{e}^{\mathrm{i}\langle u,X\rangle})$ its characteristic function. The following assertions are equivalent*:

(i) *For each orthogonal linear map $O\colon \mathbb{R}^n \to \mathbb{R}^n$ one has $OX \sim X$.*
(ii) *There is a function $\psi_X\colon \mathbb{R}^+ \to \mathbb{R}$ with $\phi_X(u) = \psi_X(\|u\|^2)$.*
(iii) *For each $a \in \mathbb{R}^n$ we have $\langle a, X\rangle \sim \|a\| X_1$, where X_1 is the first component of the vector X.*

Proof "(i)⇒(ii)": For each orthogonal linear Map O and each $u \in \mathbb{R}^n$ we have

$$\phi_X(u) = \phi_{OX}(u) = \mathrm{E}\big(\mathrm{e}^{\mathrm{i}\langle u,OX\rangle}\big) = \mathrm{E}\big(\mathrm{e}^{\mathrm{i}\langle O^\top u,X\rangle}\big) = \phi_X\big(O^\top u\big)$$

The characteristic function $\phi_X(\cdot)$ is therefore invariant under orthogonal transformations and the property (ii) follows.

"(ii)⇒(iii)": Assume $a \in \mathbb{R}^n$. Then we get for each $t \in \mathbb{R}$

$$\phi_{\langle a,X\rangle}(t) = E\big(\mathrm{e}^{\mathrm{i}t\langle a,X\rangle}\big) = \mathrm{E}\big(\mathrm{e}^{\mathrm{i}\langle ta,X\rangle}\big) = \phi_X(ta) = \psi_X\big(t^2\|a\|^2\big).$$

On the other hand, we have

$$\phi_{\|a\|X_1}(t) = \mathrm{E}\big(\mathrm{e}^{\mathrm{i}t\|a\|X_1}\big) = \mathrm{E}\big(\mathrm{e}^{\mathrm{i}\langle t\|a\|\mathbf{e}_1,X\rangle}\big) = \phi_X\big(t\|a\|\mathbf{e}_1\big) = \psi_X\big(t^2\|a\|^2\big),$$

and the property (iii) follows from the uniqueness of the characteristic function.

"(iii)⇒(i)": By the uniqueness of the characteristic function it is enough to show that the characteristic function of X is invariant under orthogonal transformations O. We have

$$\begin{aligned}\phi_{OX}(u) &= \mathrm{E}\big(\mathrm{e}^{\mathrm{i}\langle u,OX\rangle}\big) = \mathrm{E}\big(\mathrm{e}^{\mathrm{i}\langle O^\top u,X\rangle}\big) = \phi_{\langle O^\top u,X\rangle}(1) = \phi_{\|O^\top u\|X_1}(1)\\ &= \phi_{\|u\|X_1}(1) = \phi_{\langle u,X\rangle}(1) = \mathrm{E}\big(\mathrm{e}^{\mathrm{i}\langle u,X\rangle}\big) = \phi_X(u).\end{aligned}$$ □

Lemma 2.8 *The risk measure* VaR_α *is translation invariant, positive homogeneous and monotone.*

Proof Assume $a \in \mathbb{R}$ and that X, Y are arbitrary random variables.

Translation invariance: Obviously we have $F_{X+a}(x) = \mathbf{P}(X + a \le x) = \mathbf{P}(X \le x - a) = F_X(x - a)$. It follows that

$$\begin{aligned}\mathrm{VaR}_\alpha(X + a) &= \inf\big\{x : F_{X+a}(x) \ge \alpha\big\} = \inf\big\{x : F_X(x - a) \ge \alpha\big\}\\ &= \inf\big\{x + a : F_X(x) \ge \alpha\big\} = a + \inf\big\{x : F_X(x) \ge \alpha\big\}\\ &= \mathrm{VaR}_\alpha(X) + a.\end{aligned}$$

Positive Homogeneity: For $a = 0$ positive homogeneity holds trivially. If $a > 0$, then $F_{aX}(x) = \mathbf{P}(aX \le x) = \mathbf{P}(X \le \frac{x}{a}) = F_X(\frac{x}{a})$. Thus it follows that

$$\mathrm{VaR}_\alpha(aX) = \inf\big\{x : F_{aX}(x) \ge \alpha\big\} = \inf\left\{x : F_X\left(\frac{x}{a}\right) \ge \alpha\right\}$$

$$
\begin{aligned}
&= \inf\{ax : F_X(x) \geq \alpha\} = a \inf\{x : F_X(x) \geq \alpha\} \\
&= a \operatorname{VaR}_\alpha(X).
\end{aligned}
$$

Monotony: Suppose $X \geq Y$ almost everywhere. Then $F_X(x) = \mathbf{P}(X \leq x) \leq \mathbf{P}(Y \leq x) = F_Y(x)$ and so $\{x : F_X(x) \geq \alpha\} \subseteq \{x : F_Y(x) \geq \alpha\}$. We deduce that

$$
\operatorname{VaR}_\alpha(X) = \inf\{x : F_X(x) \geq \alpha\} \geq \inf\{x : F_Y(x) \geq \alpha\} = \operatorname{VaR}_\alpha(Y). \qquad \square
$$

Theorem 2.2 *When restricted to a vector space of normally distributed random variables the risk measure* $\operatorname{VaR}_\alpha$ *is coherent for any* $\alpha \in \,]\frac{1}{2}, 1[$.

Proof By Lemma 2.8 we only need to show subadditivity. Suppose $X, Y : \Omega \to \mathbb{R}$ are arbitrary normal random variables from the vector space. Because of the vector space property any linear combination of X and Y is normally distributed, so that the vector (X, Y) is multivariately normally distributed. Therefore there exist a two-dimensional random vector $Z = (Z_1, Z_2)$ which is standard normally distributed, a linear map $A : \mathbb{R}^2 \to \mathbb{R}^2$ and a vector $b = (b_1, b_2) \in \mathbb{R}^2$, so that $(X, Y)^\top = AZ + b$ holds. Since $\phi_Z(u) = \mathrm{e}^{-\|u\|^2/2}$ by Lemma 2.7 for each vector $a \in \mathbb{R}^2$ the relation $\langle a, Z\rangle \sim \|a\| Z_1$ holds. We have

$$
\begin{aligned}
X - b_1 &= \langle A^\top \mathbf{e}_1, Z\rangle \sim \|A^\top \mathbf{e}_1\| Z_1, \\
Y - b_2 &= \langle A^\top \mathbf{e}_2, Z\rangle \sim \|A^\top \mathbf{e}_2\| Z_1, \\
X + Y - b_1 - b_2 &= \langle A^\top \mathbf{e}_1 + A^\top \mathbf{e}_2, Z\rangle \sim \|A^\top \mathbf{e}_1 + A^\top \mathbf{e}_2\| Z_1.
\end{aligned}
$$

Thus by translation invariance and positive homogeneity of $\operatorname{VaR}_\alpha$ we have

$$
\begin{aligned}
\operatorname{VaR}_\alpha(X) &= \|A^\top \mathbf{e}_1\| \operatorname{VaR}_\alpha(Z_1) + b_1, \\
\operatorname{VaR}_\alpha(Y) &= \|A^\top \mathbf{e}_2\| \operatorname{VaR}_\alpha(Z_1) + b_2, \\
\operatorname{VaR}_\alpha(X + Y) &= \|A^\top \mathbf{e}_1 + A^\top \mathbf{e}_2\| \operatorname{VaR}_\alpha(Z_1) + b_1 + b_2.
\end{aligned}
$$

The subadditivity now follows from

$$
\|A^\top \mathbf{e}_1 + A^\top \mathbf{e}_2\| \leq \|A^\top \mathbf{e}_1\| + \|A^\top \mathbf{e}_2\|
$$

and $\operatorname{VaR}_\alpha(Z_1) \geq 0$ for $\alpha \geq \frac{1}{2}$, since Z_1 has a standard normal distribution. $\square$

Remark 2.5 A random variable that satisfies one of the equivalent conditions of Lemma 2.7 is called *spherical*. An affine transform of a spherical random variable is called *elliptical*. In the proof of Theorem 2.2 we only used the property of the normal distribution that multi-normal distributions are elliptic. The theorem can thus be generalized to distributions that can be written as linear combinations of components of elliptic distributions. For a precise formulation see [6, Theorem 6.8].

Theorem 2.3 *The expected shortfall* ES_α *is coherent.*

Proof Let $\mathrm{d}p$ be the Lebesgue measure. Then translation invariance, positive homogeneity and monotony follow directly from Theorem 2.1 and Lemma 2.8 with $\rho_p = \mathrm{VaR}_p$ and $(A, \mathscr{A}, \mu) = ([\alpha, 1], \mathscr{B}, \frac{1}{1-\alpha}\,\mathrm{d}p)$.

It remains to show subadditivity. For arbitrary random variables X, Y we obtain using Lemma 2.5(iii)

$$\begin{aligned}
&(1-\alpha)\big(\mathrm{ES}_\alpha(X) + \mathrm{ES}_\alpha(Y) - \mathrm{ES}_\alpha(X+Y)\big)\\
&\quad = \mathrm{E}\big(X 1_{X,\mathrm{VaR}_\alpha(X),\alpha} + Y 1_{Y,\mathrm{VaR}_\alpha(Y),\alpha} - (X+Y)1_{(X+Y),\mathrm{VaR}_\alpha(X+Y),\alpha}\big)\\
&\quad = \mathrm{E}\big(X(1_{X,\mathrm{VaR}_\alpha(X),\alpha} - 1_{(X+Y),\mathrm{VaR}_\alpha(X+Y),\alpha})\big)\\
&\qquad + \mathrm{E}\big(Y(1_{Y,\mathrm{VaR}_\alpha(Y),\alpha} - 1_{(X+Y),\mathrm{VaR}_\alpha(X+Y),\alpha})\big).
\end{aligned}$$

We now consider the expression $\mathrm{E}(X(1_{X,\mathrm{VaR}_\alpha(X),\alpha} - 1_{(X+Y),\mathrm{VaR}_\alpha(X+Y),\alpha}))$. By the construction of $1_{X,x,\alpha}$ we have for $X(\omega) < x$ the equation $1_{X,x,\alpha}(\omega) = 0$ and for $X(\omega) > x$ the equation $1_{X,x,\alpha}(\omega) = 1$. Since by Lemma 2.5(i) the inequality

$$0 \le 1_{X+Y,\mathrm{VaR}_\alpha(X+Y),\alpha} \le 1$$

holds, we obtain

$$1_{X,\mathrm{VaR}_\alpha(X),\alpha} - 1_{(X+Y),\mathrm{VaR}_\alpha(X+Y),\alpha} \begin{cases} \le 0, & \text{if } X(\omega) < \mathrm{VaR}_\alpha(X) \\ \ge 0, & \text{if } X(\omega) > \mathrm{VaR}_\alpha(X). \end{cases}$$

From this we see in both cases (and trivially also for $X = \mathrm{VaR}_\alpha(X)$) the inequality

$$\begin{aligned}
&X(1_{X,\mathrm{VaR}_\alpha(X),\alpha} - 1_{(X+Y),\mathrm{VaR}_\alpha(X+Y),\alpha})\\
&\quad \ge \mathrm{VaR}_\alpha(X)(1_{X,\mathrm{VaR}_\alpha(X),\alpha} - 1_{(X+Y),\mathrm{VaR}_\alpha(X+Y),\alpha}).
\end{aligned}$$

Lemma 2.5(ii) now implies

$$\begin{aligned}
&\mathrm{E}\big(X(1_{X,\mathrm{VaR}_\alpha(X),\alpha} - 1_{(X+Y),\mathrm{VaR}_\alpha(X+Y),\alpha})\big)\\
&\quad \ge \mathrm{VaR}_\alpha(X)\,\mathrm{E}\big((1_{X,\mathrm{VaR}_\alpha(X),\alpha} - 1_{(X+Y),\mathrm{VaR}_\alpha(X+Y),\alpha})\big)\\
&\quad = \mathrm{VaR}_\alpha(X)\big((1-\alpha) - (1-\alpha)\big) = 0.
\end{aligned}$$

The same argument also implies

$$\mathrm{E}\big(Y(1_{Y,\mathrm{VaR}_\alpha(Y),\alpha} - 1_{(X+Y),\mathrm{VaR}_\alpha(X+Y),\alpha})\big) \ge 0.$$

Therefore, summing up, we obtain

$$(1-\alpha)\big(\mathrm{ES}_\alpha(X) + \mathrm{ES}_\alpha(Y) - \mathrm{ES}_\alpha(X+Y)\big) \ge 0 + 0 = 0. \qquad \square$$

Theorem 2.4 *A spectral measure M_ϕ is coherent, if the weight function ϕ is (almost everywhere) monotone increasing.*

Proof Since ϕ is monotone increasing, we can define a measure on $([0,1], \mathscr{B})$ by $\phi(p) =: \nu([0, p])$. By Fubini's Theorem it follows that

$$\begin{aligned}
M_\phi(X) &= \int_0^1 \mathrm{VaR}_p(X)\phi(p)\,\mathrm{d}p = \int_0^1 \mathrm{VaR}_p(X)\left(\int_0^p \mathrm{d}\nu(\alpha)\right)\mathrm{d}p \\
&= \int_0^1 \left(\int_0^1 1_{[0,p]}(\alpha)\,\mathrm{VaR}_p(X)\,\mathrm{d}\nu(\alpha)\right)\mathrm{d}p \\
&= \int_0^1 \left(\int_0^1 1_{[\alpha,1]}(p)\,\mathrm{VaR}_p(X)\,\mathrm{d}\nu(\alpha)\right)\mathrm{d}p \\
&= \int_0^1 \left(\int_0^1 1_{[\alpha,1]}(p)\,\mathrm{VaR}_p(X)\,\mathrm{d}p\right)\mathrm{d}\nu(\alpha) = \int_0^1 \left(\int_\alpha^1 \mathrm{VaR}_p(X)\,\mathrm{d}p\right)\mathrm{d}\nu(\alpha) \\
&= \int_0^1 (1-\alpha)\,\mathrm{ES}_\alpha(X)\,\mathrm{d}\nu(\alpha),
\end{aligned}$$

where we have used the identity $1_{[0,p]}(\alpha) = 1_{[\alpha,1]}(p)$ for $\alpha, p \in [0,1]$. The assertion now follows from Theorem 2.1 with $\mathrm{d}\mu(\alpha) = (1-\alpha)\,\mathrm{d}\nu(\alpha)$, since

$$\begin{aligned}
\int_0^1 \mathrm{d}\mu(\alpha) &= \int_0^1 (1-\alpha)\,\mathrm{d}\nu(\alpha) = \int_0^1 \left(\int_\alpha^1 \mathrm{d}p\right)\mathrm{d}\nu(\alpha) \\
&= \int_0^1 \left(\int_0^1 1_{[\alpha,1]}(p)\,\mathrm{d}p\right)\mathrm{d}\nu(\alpha) = \int_0^1 \left(\int_0^1 1_{[\alpha,1]}(p)\,\mathrm{d}\nu(\alpha)\right)\mathrm{d}p \\
&= \int_0^1 \left(\int_0^1 1_{[0,p]}(\alpha)\,\mathrm{d}\nu(\alpha)\right)\mathrm{d}p = \int_0^1 \nu\big([0,p]\big)\,\mathrm{d}p = \int_0^1 \phi(p)\,\mathrm{d}p \\
&= 1.
\end{aligned}$$

□

A spectral measure is thus coherent exactly when the individual risk aversion puts higher weights on the higher losses.

2.3.2 Practical Considerations

Some risk measures such as VaR_α or $\mathrm{TailVaR}_\alpha$ are defined with a confidence level α. This confidence level permits a first intuitive impression of the desired safety level. It is necessary however to exercise a certain caution, since the safety level depends both on the risk measure and on the time horizon being considered. For example, we have seen above that a tail value at risk for a confidence level α always offers a greater safety level than a value at risk with the same confidence level α. In addition, given a confidence level α, it is clear that the safety level that can be attained increases with the time horizon under consideration.

A further important requirement is the practicability of a risk measure.

- If one knows the class of the distribution, then the problem of determining the risk reduces to estimating the parameters of the given distribution. But even if the distributions of the individual component risks are known, the aggregation to an overall distribution already brings considerable numerical problems, even in the simplest case where the random variables are independent. Therefore in practice one usually calculates the overall distribution using Monte Carlo simulations.
- Variance-reducing techniques can be employed to bring down the number of scenarios that are needed. Furthermore an approximate portfolio evaluation can reduce the numerical burden.
- If we assume that the risk distribution is determined by Monte Carlo simulations, then VaR_α and spectral measures can be calculated with similar effort. If the risk has to be more carefully studied then spectral measures have advantages, since they are defined through an integration and thus may be more stable. In Sect. 5.2.4 we will give more detail about this property using the example of a particularly intuitive allocation scheme for risk capital.

The result of the risk measure $\rho\colon \mathscr{M}(\Omega,\mathbb{R}) \to \mathbb{R}$ is itself no random variable, but like the expected value is a deterministic value. In Monte Carlo Simulations these deterministic quantities are approximated by an estimator, i.e., a random variable $R_k^{\rho,X}$ on the basis of k independent realizations of X. Here "approximation" means that for a given small bound $\varepsilon > 0$ and a given "meta-confidence level" $\tilde{\alpha}$ we have the inequality

$$\mathbf{P}\big(\big|\rho(X) - R_k^{\rho,X}\big| > \varepsilon\big) < 1 - \tilde{\alpha}. \tag{2.3}$$

The theoretical background is provided by the Weak Law of Large Numbers.

Example 2.3 Let $\rho = \mathrm{VaR}_\alpha$. In order to estimate $\mathrm{VaR}_\alpha(X)$ in a numerically stable way, we must choose a large enough number k of scenarios such that sufficiently many scenarios show a loss higher than $\mathrm{VaR}_\alpha(X)$. For example, to get more than 100 scenarios with higher losses we choose $k \in \mathbb{N}$ so large that $(1-\alpha)k > 100$. We shall use the notation

$$\mathrm{MAX}_m\big(\{a_1,\ldots,a_k\}\big)$$

for the m-highest value in the set $\{a_1,\ldots,a_k\}$. Now we can set

$$R_k^{\mathrm{VaR}_\alpha,X}(X_1,\ldots,X_k) = \mathrm{MAX}_{\lfloor(1-\alpha)k+1\rfloor}\big(\{X_1,\ldots,X_k\}\big),$$

where $\lfloor a \rfloor$ is the integer part of the real number a. For given $\varepsilon, \tilde{\alpha}$ we now choose k so large that the inequality (2.3) holds. That such a choice is possible follows intuitively from the definition of value at risk and the law of large numbers. In practice, one would not offer a proof but simply choose k so large that for successive Monte Carlo simulations the values of $R_k^{\rho,X}$ hardly differ from one another.

The number of simulations is often pragmatically determined by the available computing capacity and acceptable run times. This can lead to results which are actually not stable. In particular, if X is a heavy-tailed distribution (e.g., a Pareto distribution) then more than 100,000 simulations can easily be necessary to obtain stable results for $\mathrm{VaR}_{99.5\,\%}(X)$.

One can easily see that the estimate $R_k^{\mathrm{VaR}_\alpha, X}(X_1, \ldots, X_k)$ coincides with the value at risk of the empirical distribution function F_k of the sample values; this is because

$$F_k\big(R_k^{\mathrm{VaR}_\alpha, X}\big) = 1 - \frac{\lfloor (1-\alpha)k \rfloor}{k} \in [\alpha, \alpha + 1/k[.$$

By the Glivenko-Cantelli Theorem the empirical distribution functions F_k converge uniformly to the distribution F of X, so that here the value at risk of the empirical distribution can be used as an approximation to the value at risk of the theoretical distribution.

Example 2.4 Let $\rho = \mathrm{ES}_\alpha$. We set

$$R_k^{\mathrm{ES}_\alpha, X}(X_1, \ldots, X_k) = \frac{\sum_{m=1}^{\lfloor (1-\alpha)k \rfloor} \mathrm{MAX}_m(\{X_1, \ldots, X_k\})}{\lfloor (1-\alpha)k \rfloor}$$

and otherwise proceed as in the Example 2.3.

The theoretical background is provided by the following version of the Law of Large Numbers [9].

Theorem 2.5 *For a sequence $(X_k)_{k \in \mathbb{N}}$ of integrable i.i.d. random quantities on the probability space $(\Omega, \mathbf{P})$ the equation*

$$\lim_{k \to \infty} \frac{\sum_{m=1}^{\lfloor (1-\alpha)k \rfloor} \mathrm{MAX}_m(X_1, \ldots, X_k)}{\lfloor (1-\alpha)k \rfloor} = \mathrm{ES}_\alpha(X_1)$$

holds almost surely, where $\lfloor\ \rfloor$ denotes the integer part.

Proof Let F be the distribution of X_1. Then

$$y \mapsto \mathrm{VaR}_y(X_1) = \inf\big\{x : F(x) \geq y\big\}$$

is integrable, since by Lemma 2.1

$$\int_0^1 \big|\mathrm{VaR}_y(X_1)\big|\,\mathrm{d}y = \int_0^1 \big|\mathrm{VaR}_y(X_1)\big|\,\mathrm{d}F\big(\mathrm{VaR}_y(X_1)\big) = \int_{-\infty}^{\infty} |x|\,\mathrm{d}F(x) < \infty.$$

We put $U_i := F(X_i)$, $i = 1, \ldots, k$. Since $\mathbf{P}(\mathrm{VaR}_{U_i}(X_i) = X_i) = 1$ by Lemma 2.2, the X_i are identically distributed and $t \mapsto \mathrm{VaR}_t(X)$ monotone increasing, we have

$$\mathrm{MAX}_m(X_1, \ldots, X_k) = \mathrm{MAX}_m\big(\mathrm{VaR}_{F_{X_1}}(X_1), \ldots, \mathrm{VaR}_{F_{X_k}}(X_k)\big)$$

$$
\begin{aligned}
&= \mathrm{MAX}_m\big(\mathrm{VaR}_{F_{X_1}}(X_1), \dots, \mathrm{VaR}_{F_{X_k}}(X_1)\big) \\
&= \mathrm{VaR}_{\mathrm{MAX}_m(F(X_1),\dots,F(X_k))}(X_1) \\
&= \mathrm{VaR}_{\mathrm{MAX}_m(U_1,\dots,U_k)}(X_1) \quad \text{a.s.}
\end{aligned}
$$

Therefore it is enough to show

$$
\lim_{k\to\infty} \frac{\sum_{m=1}^{\lfloor(1-\alpha)k\rfloor} \mathrm{VaR}_{\mathrm{MAX}_m(U_1,\dots,U_k)}(X_1)}{\lfloor(1-\alpha)k\rfloor} = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_y(X_1)\,\mathrm{d}y \qquad \text{a.s.}
$$

We shall show more generally that for each integrable function $g\colon\,]0,1[\to\mathbb{R}$ the relationship

$$
\lim_{k\to\infty} \frac{\sum_{m=1}^{\lfloor(1-\alpha)k\rfloor} g(\mathrm{MAX}_m(U_1,\dots,U_k))}{\lfloor(1-\alpha)k\rfloor} = \frac{1}{1-\alpha}\int_\alpha^1 g(x)\,\mathrm{d}x \qquad \text{a.s.}
$$

holds. For this we define piecewise constant maps

$$
g_k\colon\,]0,1[\,\times\Omega\to\mathbb{R},
$$

$k\in\mathbb{N}$, by

$$
g_k(t) := g\big(\mathrm{MAX}_{\lfloor(1-t)k\rfloor+1}(U_1,\dots,U_k)\big).
$$

It follows that

$$
\int_{(\lfloor\alpha k\rfloor+1)/k}^1 g_k(t)\,\mathrm{d}t = \sum_{m=1}^{\lfloor(1-\alpha)k\rfloor} g\big(\mathrm{MAX}_m(U_1,\dots,U_k)\big).
$$

With the notation

$$
J_k(t) = \begin{cases} 0 & \text{for } 0\le t\le \frac{\lfloor\alpha k\rfloor+1}{k}, \\ \frac{k}{\lfloor(1-\alpha)k\rfloor} & \text{for } \frac{\lfloor\alpha k\rfloor+1}{k} < t\le 1, \end{cases}
$$

it is thus sufficient to show

$$
\lim_{k\to\infty}\int_0^1 g_k(t)J_k(t)\,\mathrm{d}t = \frac{1}{1-\alpha}\int_\alpha^1 g(t)\,\mathrm{d}t \quad \text{a.s.} \tag{2.4}
$$

Let λ be the Lebesgue measure on $]0,1[$. We next show that with probability 1 with respect to $(\Omega,\mathbf{P})$ we have

$$
\lim_{k\to\infty}\lambda\big(\big\{t : \big|g_k(t)-g(t)\big|\ge\delta\big\}\big) = 0 \quad \forall\delta>0. \tag{2.5}
$$

For $\varepsilon>0$ we can find by Lusin's Theorem a Borel set $B\subseteq\,]0,1[$ and a continuous function $\tilde g\colon\,]0,1[\to\mathbb{R}$, so that $g=\tilde g$ on $]0,1[\setminus B$ and $\lambda(B)\le\varepsilon$. We now put

$$\tilde{g}_k(t) := \tilde{g}\big(\mathrm{MAX}_{\lfloor(1-t)k\rfloor+1}(U_1, \ldots, U_k)\big),$$
$$B_k := \big\{t\colon\ \mathrm{MAX}_{\lfloor(1-t)k\rfloor+1}(U_1, \ldots, U_k) \in B\big\}.$$

$\tilde{g}_k$ is correspondingly piecewise constant, and we have $\{t\colon \tilde{g}_k(t) \neq g_k(t)\} \subseteq B_k$. Since the U_i are identically distributed and independent,

$$\lambda(B_k) = \frac{1}{k}\sum_{i=1}^{k} 1_B(U_i)$$

converges a.s., according to the Strong Law of Large Numbers, to

$$\mathrm{E}\big(1_B(U_1)\big) = \mathbf{P}(U_1 \in B) = \lambda(B) \leq \varepsilon,$$

so, in particular, $\limsup_k \lambda(B_k) \leq \varepsilon$ a.s. Since $\mathrm{MAX}_{\lfloor(1-t)k\rfloor+1}(U_1, \ldots, U_k)$, as the $\frac{\lfloor tk\rfloor+1}{k}$-quantile of the empirical distribution of the samples $(U_1, \ldots, U_k)$, converges to the t-quantile of the uniform distribution and $\tilde{g}$ is continuous, it also holds that

$$\lim_{k\to\infty} \tilde{g}_k(t) = \tilde{g} \quad \text{a.s.}$$

All in all we conclude that

$$\begin{aligned}
\limsup_k \lambda\big(\big\{t : \big|g_k(t) - g(t)\big| \geq \delta\big\}\big) &\leq \limsup_k \lambda\big(\big\{t : \big|\tilde{g}(t) - g(t)\big| \geq \delta\big\}\big)\\
&\quad + \limsup_k \lambda\big(\big\{t : \big|g_k(t) - \tilde{g}_k(t)\big| \geq \delta\big\}\big)\\
&\quad + \limsup_k \lambda\big(\big\{t : \big|\tilde{g}_k(t) - \tilde{g}(t)\big| \geq \delta\big\}\big)\\
&\leq \lambda(B) + \limsup_k \lambda(B_k)\\
&\quad + \limsup_k \lambda\big(\big\{t : \big|\tilde{g}_k(t) - \tilde{g}(t)\big| \geq \delta\big\}\big)\\
&\leq 2\varepsilon.
\end{aligned}$$

With this the relation (2.5) has been shown.

Since we also have

$$\lim_{k\to\infty}\int_0^1 |g_k|\,\mathrm{d}\lambda = \lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}\big|g(U_i)\big| = \int_0^1 |g|\,\mathrm{d}\lambda,$$

we can for almost every $\omega \in \Omega$ apply the Theorem of Vitali relative to $(]0, 1[, \lambda)$ to obtain

$$\lim_{k\to\infty}\int_0^1 |g_k - g|\,\mathrm{d}\lambda = 0 \quad \text{a.s.}$$

Since the sequence J_k, $k \in \mathbb{N}$, is bounded and converges to $\frac{1}{1-\alpha}\mathbf{1}_{(\alpha,1)}$, we finally obtain the desired convergence (2.4). □

2.4 Dynamic Risk Measures

The risk measures that we have so far studied are as a rule intended for a time horizon of one year. On the other hand, insurance contracts and the liabilities associated to them are often at risk for many years. This temporal asymmetry raises the following questions:

- How should a risk measure reflect new information that becomes available with the passage of time?
- How should a risk measure react to the changes in the risk profile over a multi-year time horizon?
- How should one take account of temporal dependencies?

Temporal dependencies can be induced by external trends that are relevant to the development of losses. An example from life insurance is the improvement in life expectancy that results from medical progress. The nature of the loss to be insured can also change with time. For example, older persons have a greater probability of dying than younger ones, and the associated volatility is correspondingly greater. Therefore life insurances have a risk profile that changes with time. This can have consequences for the risk capital needed.

Example 2.5 A company takes over at the time $t = 0$ the obligations of a competitor for a purchase price of V_0. The inventory is exhausted in n years. The company expects that in year t it should have reserves V_t (deterministically calculated at time 0). Furthermore suppose the insurance benefit L_t in year t should follow a standard normal distribution with expectation μ_t and standard deviation σ_t.

The cash flow at time t is then given by the equation

$$\mathrm{Cf}_t = (1 + s_t)V_{t-1} - V_t - L_t,$$

where we have denoted the risk-free interest rate (assumed deterministic) by s_t. With the notation

$$v_t = \prod_{\tau=1}^{t} (1 + s_\tau)^{-1}$$

for the discount factor, the present value of the cash flow is given by

$$W_1 = \sum_{t=1}^{n} v_t \mathrm{Cf}_t = V_0 - \sum_{t=1}^{n} v_t L_t.$$

Obviously W_1 is also normally distributed, and we have

$$\mathrm{E}(W_1) = V_0 - \sum_{t=1}^{n} v_t \mu_t.$$

In this the index $_1$ is for the start of the first time period (see Fig. 2.4).

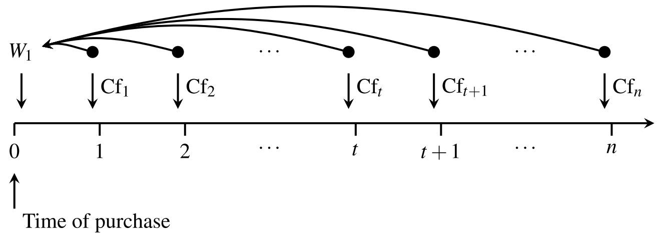

Fig. 2.4 The value W_1 of the portfolio purchased

The random vector $(L_1, \ldots, L_n)^\top$ has the covariance matrix

$$\mathrm{cov}\big((L_1, \ldots, L_n)^\top\big)_{ij} = \mathrm{corr}_{ij}\sigma_i\sigma_j,$$

where we have set $\mathrm{corr}(L_s, L_t) = \mathrm{corr}_{st}$. From

$$\mathrm{cov}\left(\sum_{t=1}^{n} v_t L_t\right) = (v_1, \ldots, v_t)\,\mathrm{cov}\big((L_1, \ldots, L_n)^\top\big)(v_1, \ldots, v_t)^\top$$

we obtain

$$\sigma(W_1) = \sqrt{\sum_{i,j=1}^{n} \mathrm{corr}_{ij} v_i \sigma_i v_j \sigma_j}$$

and with Proposition 2.5

$$\mathrm{ES}_\alpha(W_1) = -V_0 + \sum_{t=1}^{n} v_t \mu_t + \frac{\varphi_{0,1}(\Phi_{0,1}^{-1}(\alpha))}{1-\alpha} \sqrt{\sum_{i,j=1}^{n} \mathrm{corr}_{ij} v_i \sigma_i v_j \sigma_j},$$

where the time horizon stretches over n periods. The time dependence of the insurance benefit L_t increases the risk and thus the required risk capital, since for $\mathrm{corr}_{ij} > 0$ the inequality

$$\sum_{i,j=1}^{n} \mathrm{corr}_{ij} v_i \sigma_i v_j \sigma_j - \sum_{i,j=1}^{n} (v_i \sigma_i)^2 = 2 \sum_{i<j}^{n} \mathrm{corr}_{ij} v_i \sigma_i v_j \sigma_j > 0$$

holds.

This example shows that it is not possible to describe the risk over several periods purely based on the corresponding one-period risks. If we want to capture correctly multi-year risks we must consider multi-period risk measures. For risk management purposes it is thus of interest to describe the changes in risks with the passage of time. We are therefore interested not only in $\mathrm{ES}_\alpha(W_1)$ but also in $\mathrm{ES}_\alpha(W_t)$, where $\mathrm{ES}_\alpha(W_t)$is determined for the time horizon $n - t + 1$. At time $t = 0$, $\mathrm{ES}_\alpha(W_t)$ is a random variable, since the evolution of losses for the first $t-1$ periods is still unknown.

In Sect. 2.4.2 we will present a framework for the description of these dynamic aspects. As preparation we need some fundamental facts about filtrations.

2.4.1 Filtrations

In this section we introduce terminology that allows us to describe the interaction of time and known information.

Definition 2.9 Let $\mathbb{T} = \{0, \dots, n\}$ and $\mathscr{F}_0 = \{\emptyset, \Omega\}$. A *filtration* $(\mathscr{F}_t)_{t\in\mathbb{T}}$ is a set of σ-algebras on Ω with

$$\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots \subseteq \mathscr{F}_n.$$

Remark 2.6 The condition that $\mathscr{F}_0$ is the trivial σ-algebra, is often not required. We do require it here to express the fact that the initial values can be considered known. On technical grounds, it is frequently required that $\mathscr{F}_0$ be the σ-algebra generated by the null sets with respect to a given measure. This extra property can always be achieved by completing the σ-algebra, and we don't need it for what follows.

Remark 2.7 Definition 2.9 can also be generalized to infinite sets $\mathbb{T} \subseteq \mathbb{N}$ and continuous index sets $\mathbb{T} \subseteq \mathbb{R}$. For our purposes, however, finitely many discrete time steps are sufficient.

Definition 2.10 Let $(\mathscr{F}_t)_{t\in\mathbb{T}}$ be a filtration on the set Ω. An *adapted stochastic process with values in* $\mathbb{R}^k$ is a map

$$X : \Omega \times \mathbb{T} \to \mathbb{R}^k, \quad (\omega, t) \mapsto X_t(\omega),$$

where for each $t \in \mathbb{T}$ the map $X_t : \Omega \to \mathbb{R}^k$ is measurable with respect to $\mathscr{F}_t$.

This definition can be interpreted as follows: Each $\omega \in \Omega$ describes a possible history of the process under consideration. At time t the part of the history that corresponds to the past $\{0, \dots, t\}$ is known. The adaptedness condition means that the value $X_t(\omega)$ at time t is known with certainty. This interpretation is particularly striking for the product filtration (Sect. 2.4.3), see Corollary 2.2.

Example 2.6 (Finitely generated filtration) The σ-algebras that occur in practical modeling are mostly finitely generated, because computers are used. Let $\mathscr{P}_1 = \{A_1, \dots, A_{m_1}\}$ be a finite *partition* of Ω. In other words, suppose $A_j \subset \Omega$, $A_i \cap A_j = \emptyset$ for $i \neq j$ and $\bigcup_{i=1}^{m_1} A_i = \Omega$. Let $\mathscr{F}_1$ be the σ-algebra generated by $\mathscr{P}_1$. Now we can construct inductively a filtration as successive refinements of $\mathscr{P}_1$. For each $A \in \mathscr{P}_t$ we choose a partition $\mathscr{P}(A) = \{B_1(A), \dots, B_{k(A)}(A)\}$ and set

$\mathscr{P}_{t+1} = \bigcup_{A\in\mathscr{P}_t} \mathscr{P}(A)$. This family of subsets of Ω is then again a partition of Ω and thus defines a σ-algebra $\mathscr{F}_{t+1}$. A concrete example is given by

$$\begin{bmatrix} \mathscr{P}_0 = \{\{1,2,3,4,5\}\} \\ \mathscr{P}_1 = \{\{1,2\},\{3,4,5\}\} \\ \mathscr{P}_2 = \{\{1,2\},\{3\},\{4,5\}\} \end{bmatrix} \to \begin{bmatrix} \mathscr{F}_0 = \{\emptyset,\{1,2,3,4,5\}\} \\ \mathscr{F}_1 = \{\emptyset,\{1,2\},\{3,4,5\},\{1,2,3,4,5\}\} \\ \mathscr{F}_2 = \{\emptyset,\{3\},\{1,2\},\{4,5\},\{1,2,3\}, \\ \qquad \{3,4,5\},\{1,2,4,5\},\{1,2,3,4,5\}\} \end{bmatrix}.$$

From

$$X_t(\omega) = \begin{cases} t & \text{for } \omega = 1 \\ t & \text{for } \omega = 2 \\ 3t + 4t(t-1) & \text{for } \omega = 3 \\ 3t & \text{for } \omega = 4 \\ 3t & \text{for } \omega = 5 \end{cases} \quad \text{and} \quad Y(\omega) = \begin{cases} t & \text{for } \omega = 1 \\ t & \text{for } \omega = 2 \\ 3t + 4t^2 & \text{for } \omega = 3 \\ 3t & \text{for } \omega = 4 \\ 3t & \text{for } \omega = 5 \end{cases}$$

we see X_t is an adapted stochastic process, though Y_t is not: Obviously Y_0 and X_0 are constant and so $\mathscr{F}_0$-measurable. We have $X_1(\omega) = 1 \Leftrightarrow \omega \in \{1,2\} \in \mathscr{F}_1$, and $X_1(\omega) = 3 \Leftrightarrow \omega \in \{3,4,5\} \in \mathscr{F}_1$. Therefore X_1 is $\mathscr{F}_1$-measurable. On the other hand, $Y_1(\omega) = 7 \Leftrightarrow \omega \in \{3\} \notin \mathscr{F}_1$, so that Y_1 is not $\mathscr{F}_1$-measurable. Finally we have $X_2(\omega) = 2 \Leftrightarrow \omega \in \{1,2\} \in \mathscr{F}_2$, $X_2(\omega) = 14 \Leftrightarrow \omega \in \{3\} \in \mathscr{F}_2$, and $X_2(\omega) = 6 \Leftrightarrow \omega \in \{4,5\} \in \mathscr{F}_2$, so that X_2 is also $\mathscr{F}_2$-measurable.

Suppose X is the final result of a long-term investment. At the time of the investment the company expected the result $\mathrm{E}(X)$. In the course of time, however, this assessment will change because of economic uncertainty. This updating process can be described in terms of conditional expectations with respect to a filtration.

Definition 2.11 Let $(\Omega, \mathscr{A}, \mathbf{P})$ be a probability space, $\tilde{\mathscr{A}}$ a sub-σ-algebra of $\mathscr{A}$ and X an $\mathscr{A}$-measurable random variable. The *conditional expectation* of X with respect to $\tilde{\mathscr{A}}$ is the (a.s. uniquely determined) $\tilde{\mathscr{A}}$-measurable random variable $\mathrm{E}(X \mid \tilde{\mathscr{A}})$ with the property that for any bounded $\tilde{\mathscr{A}}$-measurable random variable Z there holds

$$\mathrm{E}(XZ) = \mathrm{E}\big(\mathrm{E}(X \mid \tilde{\mathscr{A}})Z\big).$$

For a proof that this is well defined we cite Theorem 23.4 in [5].

Lemma 2.9 *Let $(\mathscr{F}_t)_{t\in\mathbb{T}}$ be a filtration on Ω and X a $\mathscr{F}_n$-measurable random variable. Then*

$$(t,\omega) \mapsto X_t(\omega) := \mathrm{E}(X \mid \mathscr{F}_t)_{|\omega}$$

is an adapted stochastic process.

In addition, X_t satisfies the martingale property

$$\mathrm{E}(X_{t+1} \mid \mathscr{F}_t) = X_t$$

for all $t \in \{0,\dots,n-1\}$.

Proof The first assertion follows directly from the $\mathscr{F}_t$-measurability of $\mathrm{E}(X \mid \mathscr{F}_t)$.

For the second assertion, let Z be a $\mathscr{F}_t$-measurable function. Then Z is also $\mathscr{F}_{t+1}$-measurable, and

$$\mathrm{E}(X_{t+1} Z) = \mathrm{E}\big(\mathrm{E}(X \mid \mathscr{F}_{t+1})\, Z\big) = \mathrm{E}(XZ) = \mathrm{E}\big(\mathrm{E}(X \mid \mathscr{F}_t)\, Z\big) = \mathrm{E}(X_t\, Z).$$

The claim follows from the $\mathscr{F}_t$-measurability of X_t and the uniqueness of the conditional expectation. □

By the martingale property one can interpret $\mathrm{E}(X \mid \mathscr{F}_t)$ as the best estimate for the result X at time t. $\mathrm{E}(X \mid \mathscr{F}_t)$ is itself a random variable that reflects the uncertainty between the times 0 and t.

2.4.1.1 Product Filtrations

For concrete applications to cash flows, as a rule, filtrations are constructed based on successive new pieces of information. To describe this construction we need the following notation:

Definition 2.12 Let $\mathscr{A}_1, \ldots, \mathscr{A}_k$ be σ-algebras on the sets $\Omega_1, \ldots, \Omega_k$. The *product-$\sigma$-algebra* on the product set $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_k$ is then given by

$$\bigotimes_{t=1}^{k} \mathscr{A}_t = \mathscr{A}_1 \otimes \cdots \otimes \mathscr{A}_k = \sigma\big(A_1 \times \cdots \times A_k \mid A_t \in \mathscr{A}_t \text{ for } t \in \{1, \ldots .k\}\big),$$

where we have used the convention

$$A_1 \times \cdots \times A_{t-1} \times \emptyset \times A_{t+1} \times \cdots \times A_k = \emptyset.$$

In general, we have $\mathscr{A}_1 \otimes \mathscr{A}_2 \neq \{A_1 \times A_2 \mid A_1 \in \mathscr{A}_1, A_2 \in \mathscr{A}_2\}$. We have equality, however, if $\mathscr{A}_2$ is the trivial σ-algebra.

Lemma 2.10 *Let Ω_1 and Ω_2 be sets and let $\mathscr{A}_1$ be a σ-algebra on Ω_1. Then*

$$A \in \mathscr{A}_1 \otimes \{\emptyset, \Omega_2\} \quad \Leftrightarrow \quad A = A_1 \times \Omega_2,$$

with $A_1 \in \mathscr{A}_1$.

Proof This follows immediately from the fact that the sets of the form

$$\{A_1 \times \Omega_2 \colon A_1 \in \mathscr{A}_1\}$$

make up a σ-algebra. □

Lemma 2.11 (Associativity) *Assume $\mathscr{A}_1, \ldots, \mathscr{A}_{j+k}$ to be σ-algebras. Then*

$$\left(\bigotimes_{t=1}^{j} \mathscr{A}_t\right) \otimes \left(\bigotimes_{t=j+1}^{j+k} \mathscr{A}_t\right) = \bigotimes_{t=1}^{j+k} \mathscr{A}_t.$$

Proof See [3, page 142f]. □

A σ-algebra models what events are in principle possible. To describe economic dynamics we assume that, as a matter of principle, the same events are possible in each time period. For example, in each time period t the event E_t can occur that the stock price of a corporation increases by more than 10 %. On the other hand, the probability of the events depends on the dynamics, and is thus in general different for each period of time: If the corporation at the end of period $t-1$ brings a new very promising product to market, then the probability of the event E_t will be higher than the probability of E_{t-1}. In this approach we assign each time period (considered in isolation) the same σ-algebra $\mathscr{A}$ but not necessarily the same probability measure. The following example illustrates this idea:

Example 2.7 (AR(1)-Process) We consider a stock index S_t, whose dynamics are governed by the equation

$$S_t = \alpha S_{t-1} + s\omega_t, \quad t \in \{1, \ldots, n\},$$

where $\alpha, s > 0$ are constants and $\omega_1, \omega_2, \ldots$ are drawn independently from a standard normal distribution. For each fixed period t, our σ-algebra is exactly the Borel algebra $\mathscr{B}(\mathbb{R})$ on $\mathbb{R}$. To describe the whole dynamics we need to draw n times, and clearly the associated σ-algebra is exactly the Borel algebra

$$\mathscr{B}\big(\mathbb{R}^n\big) = \bigotimes_{t=1}^{n} \mathscr{B}(\mathbb{R})$$

on the set $\Omega = \mathbb{R}^n$. To describe the dynamics up to the period t, it would seem natural to choose as the σ-algebra the Borel algebra $\mathscr{B}(\mathbb{R}^t)$. However, this has the disadvantage that the set on which the σ-algebra is defined changes with each time step. If we multiply $\mathscr{B}(\mathbb{R}^t)$ together with $n-t$ copies of the trivial σ-algebra $\mathscr{A}_0 = \{\emptyset, \mathbb{R}\}$, the resulting σ-algebra is defined on the whole space $\Omega = \mathbb{R}^n$. It has moreover the same measurability structure as the σ-algebra $\mathscr{B}(\mathbb{R}^t)$. This is because, since the maps $f(\omega_1, \ldots, \omega_n)$ measurable with respect to the σ-algebra $\mathscr{B}(\mathbb{R}^t) \times \bigotimes_{s=t+1}^{n} \mathscr{A}_0$ do not depend on $(\omega_{t+1}, \ldots, \omega_n)$, there is a bijection

$$\Psi(g)(\omega_1, \ldots, \omega_n) = g(\omega_1, \ldots, \omega_t)$$

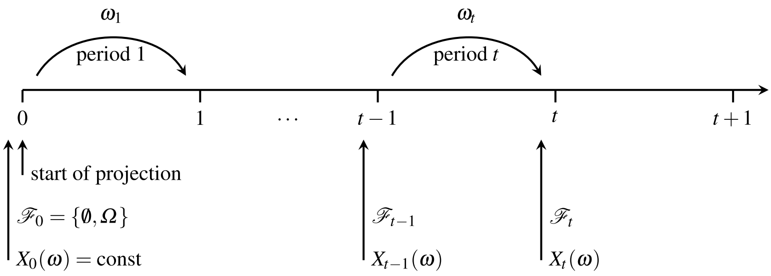

Fig. 2.5 Illustration for Definition 2.13. The random events occurring in the period t are described by the drawing $\omega_t \in \Omega_t$. If X_t is an adapted stochastic process then at time t the values $X_0(\omega), \dots, X_t(\omega)$ are known, since they depend only on $(\omega_1, \dots, \omega_t)$ (see Corollary 2.2 further below)

of the space of $\mathscr{B}(\mathbb{R}^t)$-measurable maps onto the space of $\mathscr{B}(\mathbb{R}^t) \times \bigotimes_{s=t+1}^n \mathscr{A}_0$-measurable maps. With this the natural filtration for our process is

$$\mathscr{F}_t = \bigotimes_{s=1}^{t} \mathscr{B}(\mathbb{R}) \otimes \bigotimes_{s=t+1}^{n} \mathscr{A}_0.$$

The process describing the dynamics

$$S\colon \mathbb{R}^n \times \{1, \dots, n\} \to \mathbb{R},$$
$$(\omega_1, \dots, \omega_n, t) \mapsto S_t(\omega_1, \dots, \omega_n)$$

is an adapted stochastic process for this filtration. Note that S_t does not depend on $\omega_{t+1}, \dots, \omega_n$. This expresses that the future is not known.

Remark also that for $\alpha \notin \{0, 1\}$ neither the S_t for $t \in \{1, \dots, n\}$ nor the increments $S_{t+1} - S_t$ for $t \in \{1, \dots, n-1\}$ are independently distributed. The distribution of the processes S therefore shows a non-trivial dependence structure.

The construction described in Example 2.7 can be generalized in the following way:

Definition 2.13 Assume $\mathbb{T} = \{0, \dots, n\}$ and for $t \in \mathbb{T} \setminus \{0\}$ that $\mathscr{A}_t$ is a σ-algebra on the set Ω_t. The *product filtration* on the Cartesian product $\Omega = \prod_{t=1}^n \Omega_t$ is given by

$$\mathscr{F}_t = \begin{cases} \{\emptyset, \Omega\} & \text{if } t = 0, \\ \bigotimes_{s=1}^t \mathscr{A}_s \otimes \bigotimes_{s=t+1}^n \{\emptyset, \Omega_s\} & \text{otherwise.} \end{cases} \tag{2.6}$$

The σ-algebra $\mathscr{F}_t$ can be understood as the restriction of the σ-Algebra $\mathscr{F}_n$ to the time interval from 0 to t (see the Fig. 2.5). In practical applications the $\mathscr{A}_s$ are almost always the same. The somewhat greater generality of Definition 2.13 does not lead to any additional difficulties.

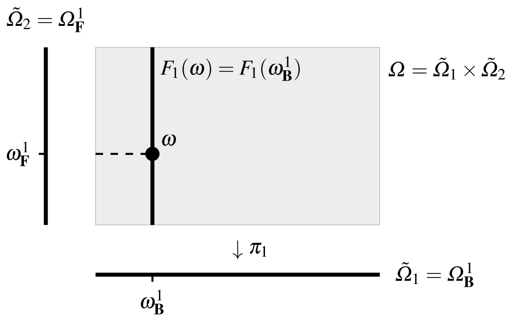

Fig. 2.6 Illustration of the product structure in Definition 2.14 in a two-dimensional example

Definition 2.14 Let $\Omega = \prod_{t=1}^n \Omega_t$ and

$$\pi_t : \Omega \to \prod_{s=1}^{t} \Omega_s, \quad \omega \mapsto \pi_t(\omega) = (\omega_1, \dots, \omega_t)$$

be the projection on the first t factors. For $\omega \in \Omega$ and $t \in \{1, \dots, n\}$ the *t-fiber through ω* is given by

$$F_t(\omega) = \pi_t^{-1}\big(\pi_t(\omega)\big),$$

and we put $F_0(\omega) = \Omega$. For $w \in \pi_t(\Omega)$ the *t-fiber over w* is the set $F_t(w) = \pi_t^{-1}(w)$.

We write $\Omega_{\mathbf{B}}^t = \prod_{s=1}^t \Omega_s$ and $\Omega_{\mathbf{F}}^t = \prod_{s=t+1}^n \Omega_s$.[4] Furthermore we use the notation $\pi_t(\omega) = \omega_{\mathbf{B}}^t \in \Omega_{\mathbf{B}}^t$ and define $\omega_{\mathbf{F}}^t \in \Omega_{\mathbf{F}}^t$ by $\omega = (\omega_{\mathbf{B}}^t, \omega_{\mathbf{F}}^t)$.

Definition 2.14 is illustrated in Fig. 2.6 and in Fig. 2.7.

Corollary 2.1 *Let $\Omega = \prod_{t=1}^n \Omega_t$. For each $\omega \in \Omega$ we thus have*

$$\{\omega\} = F_n(\omega) \subseteq F_{n-1}(\omega) \subseteq \cdots \subseteq F_1(\omega) \subseteq F_0(\omega) = \Omega.$$

The relationships in Definition 2.14 are consistent in the sense that the equation $F_t(\omega) = F_t(\pi_t(\omega))$ holds for every $\omega \in \Omega$.

The following lemma, resp. the subsequent corollary, shows that a stochastic process adapted to a product filtration depends at time t only on the uncertainty up to the time t, and not on future imponderables. This conforms to the experience that the present depends on events in the past, but is not influenced by events in the future.

Lemma 2.12 *A map $g: \Omega \to \mathbb{R}$ is exactly then $\mathscr{F}_t$-measurable when it is $\mathscr{F}_n$-measurable and constant on the fiber $F_t(\omega)$.*

---

[4]The index $_{\mathbf{B}}$ stands for "Basis" and the index $_{\mathbf{F}}$ for "Fiber".

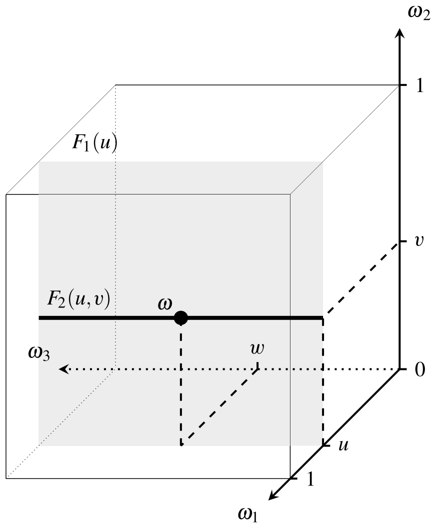

Fig. 2.7 Illustration for Definition 2.14. We consider the space $\Omega =]0, 1[^3$, where $\mathbf{P}$ is the Lebesgue measure; $\omega = (u, v, w)$ is the random value that has to be realized. At time $t = 1$ at the beginning of the period $t + 1$, $\omega_1 = u$ is known. The fiber $F_1(u)$ describes the probability space for the remaining uncertainty. At time $t = 2$, (u, v) is known and the fiber $F_2(u, v) \subseteq F_1(u)$ describes the remaining uncertainty. At time $t = 3$, ω is known. Since there is no further uncertainty, the fiber $F_3(\omega)$ is reduced to the point ω

Proof "⇒": Let g be $\mathscr{F}_t$-measurable. $\mathscr{F}_t \subseteq \mathscr{F}_n$, so g is also $\mathscr{F}_n$-measurable. From Lemma 2.10 follows that $\mathscr{F}_t$ is made up of the $\mathscr{F}_n$-measurable subsets of the form $A = \tilde{A} \times \Omega^t_{\mathbf{F}}$. We now assume that g is not constant on the fibers. Because $F_t(\omega) = \{\omega^t_{\mathbf{B}}\} \times \Omega^t_{\mathbf{F}}$, there exist $x, y \in \Omega^t_{\mathbf{F}}$ with $g(\omega^t_{\mathbf{B}}, x) \neq g(\omega^t_{\mathbf{B}}, y)$. Thus there are open intervals B_x, B_y with $g(\omega^t_{\mathbf{B}}, x) \in B_x$, $g(\omega^t_{\mathbf{B}}, y) \in B_y$ and $B_x \cap B_Y = \emptyset$. It follows that $g^{-1}(B_x) \cap g^{-1}(B_y) = \emptyset$. Since $(\omega^t_{\mathbf{B}}, x) \in g^{-1}(B_x) \setminus g^{-1}(B_y)$, we deduce that $g^{-1}(B_y) \cap F_t(\omega) \neq F_t(\omega)$. Thus we have a contradiction to $g^{-1}(B_y) \in \mathscr{F}_t$.

"⇐": Since g is constant on the fibers $F_t(\omega)$, for each Borel set $B \subset \mathbb{R}$ there exists a set $A \subset \Omega^t_{\mathbf{B}}$ with $g^{-1}(B) = A \times \Omega^t_{\mathbf{F}}$. Since the set $g^{-1}(B)$ is measurable with respect to $\mathscr{F}_n$, it is also $\mathscr{F}_t$-measurable by Lemma 2.10. □

Corollary 2.2 *Let $(\mathscr{F}_t)_{t \in \mathbb{T}}$ be a product filtration on $\Omega = \prod_{t=1}^n \Omega_t$ and $(\omega, t) \mapsto X_t(\omega)$ an adapted stochastic process. Then $X_t(\omega)$ depends only on the first t components $\omega_1, \ldots, \omega_t$.*

Proof This follows because $X_t(\omega)$ is constant on $F_t(\omega) = \{\omega^t_{\mathbf{B}}\} \times \prod_{s=t+1}^n \Omega_s$. □

Since $\mathscr{F}_t$-measurable functions g only depend on ω through $\omega^t_{\mathbf{B}}$, we shall write $g(\omega^t_{\mathbf{B}})$ in some places instead of $g(\omega)$.

Definition 2.15 For $t \in \{1, \ldots, n\}$, let (Ω_t, μ_t) be a measure space and $(\mathscr{F}_t)_{t \in \mathbb{T}}$ the product filtration on $\Omega = \prod_{t=1}^n \Omega_t$. Suppose the probability measure $\mathbf{P}$ on $(\Omega, \mathscr{F}_n)$ is absolutely continuous with respect to the product measure $\mu = \bigotimes_{t=1}^n \mu_t$. Then $(\Omega, (\mathscr{F}_t)_{t \in \mathbb{T}}, \mathbf{P})$ is a *filtered product economy*.

We write $\mu^t_{\mathbf{B}} = \bigotimes_{s=1}^t \mu_s$ and $\mu^t_{\mathbf{F}} = \bigotimes_{s=t+1}^n \mu_s$.

Remark 2.8 Since for each $A \subseteq F_t(\omega)$ there is a unique set $\tilde{A} \subseteq \prod_{s=t+1}^{n} \Omega_s$ with $A = (\omega_1, \dots, \omega_t) \times \tilde{A}$, $\mu_{\mathbf{F}}^t$ induces, in a canonical way, through

$$\mu_{\mathbf{F}}^{t,F_t(\omega)}(A) = \mu_{\mathbf{F}}^t(\tilde{A})$$

a measure on $F_t(\omega)$. To simplify the notation, we will also write $\mu_{\mathbf{F}}^t$ for this measure on the fiber.

In a filtered product economy, n discrete time periods are modeled, so that, during each period t, additional economic uncertainty arises that is described by $\omega_t \in \Omega_t$. The $\omega = (\omega_1, \dots, \omega_n)$ thus describes the cumulative uncertainty over all periods. At the start of period $t+1$, $\omega_{\mathbf{B}}^t$ is a known quantity while $F_t(\omega_{\mathbf{B}}^t)$ describes the remaining risk. An economic product economy does not describe the setting of prices, but describes only the random events that can influence fixing prices. The price dynamics of an investment are described by adapted stochastic processes on economic product economies (see Example 2.7).

Remark 2.9 The measures μ_t on $(\Omega_t, \mathscr{A}_t)$ model the sources of randomness in the individual time periods t. By choosing the product measure $\mu = \bigotimes_{t=1}^{n} \mu_t$ on $\Omega = \prod_{t=1}^{n} \Omega_t$ we are supposing the independence of the sources of randomness. Each period thus brings, independently of all other periods, a random influence ω_t to the cumulative uncertainty $\omega = (\omega_1, \dots, \omega_n)$. This is, however, just a mathematical trick to simplify the modeling, and the product measure μ has, as a rule, no direct economic interpretation. In particular, the independence of the sources of randomness does not mean that an economic price dynamics on $(\Omega, (\mathscr{F}_t)_{t\in\mathbb{T}}, \mu)$ has independent increments, as Example 2.7 shows.

As we are merely assuming that $\mathbf{P}$ is absolutely continuous relative to μ, we have defined a modeling framework that is sufficiently general for practical applications.

Example 2.8 Consider two projection periods $t \in \{1, 2\}$ and a loss function that is exponentially distributed in each period. We further assume that the economic cycle has the effect that at the end of period 1 the expected loss for period 2 is 10 % higher than the losses incurred in period 1. Let (X_1, X_2) be the loss process over the two periods. Then the corresponding distribution function $\mathbf{P}(X_1 < y_1, X_2 < y_2)$ has the density

$$p(y_1, y_2) = a \exp(-a y_1) \frac{10}{11 y_1} \exp\left(-\frac{10 y_2}{11 y_1}\right) 1_{(0,\infty)\times(0,\infty)}(y_1, y_2)$$

with respect to the two-dimensional Lebesgue measure, where $\mathrm{E}(X_1) = 1/a$. In each simulation, one draws a random number x_1 from the exponential distribution with rate a, and then a random number x_2 from an exponential distribution with rate $(1.1\, x_1)^{-1}$.

Thus, from the standpoint of practical modeling, using a probability measure $\mathbf{P}$ which is absolutely continuous relative to μ is sufficiently general, and leads to significant technical simplifications (see e.g., Lemma 2.13 and Proposition 2.7).

Lemma 2.13 *Let* $(\Omega, (\mathscr{F}_t)_{t\in\mathbb{T}}, \mathbf{P})$ *be a filtered product economy. On the fiber* $F_t(\omega)$ *there is a probability measure on the* σ*-algebra induced by* $\mathscr{F}_n$*, given by*

$$\mathbf{P}_{\omega_{\mathbf{B}}^t}(A) = \frac{\int 1_A\, p(\omega_{\mathbf{B}}^t, \omega_{\mathbf{F}}^t)\,\mathrm{d}\mu_{\mathbf{F}}^t}{\int p(\omega_{\mathbf{B}}^t, \omega_{\mathbf{F}}^t)\,\mathrm{d}\mu_{\mathbf{F}}^t},$$

where p *is the density of* $\mathbf{P}$ *relative to* μ*.*

Proof Clearly $\mathbf{P}_{\omega_{\mathbf{B}}^t}$ is a measure on $F_t(\omega) = \{\omega_1\} \times \cdots \times \{\omega_t\} \times \Omega_{t+1} \times \cdots \times \Omega_n$. The assertion thus follows from

$$\mathbf{P}_{\omega_{\mathbf{B}}^t}\big(F_t(\omega)\big) = \frac{\int 1_{F_t(\omega)}\, p(\omega_{\mathbf{B}}^t, \omega_{\mathbf{F}}^t)\,\mathrm{d}\mu_{\mathbf{F}}^t}{\int p(\omega_{\mathbf{B}}^t, \omega_{\mathbf{F}}^t)\,\mathrm{d}\mu_{\mathbf{F}}^t} = \frac{\int p(\omega_{\mathbf{B}}^t, \omega_{\mathbf{F}}^t)\,\mathrm{d}\mu_{\mathbf{F}}^t}{\int p(\omega_{\mathbf{B}}^t, \omega_{\mathbf{F}}^t)\,\mathrm{d}\mu_{\mathbf{F}}^t} = 1. \qquad \square$$

We have already seen that an observer at time t knows the portion $\omega_{\mathbf{B}}^t$ of his history ω, and that the remaining uncertainty can be described by $F_t(\omega_{\mathbf{B}}^t)$. The probability measure $\mathbf{P}_{\omega_{\mathbf{B}}^t}$ is of use to the observer in determining the probabilities of future events.

Proposition 2.7 *Let* $(\Omega, (\mathscr{F}_t)_{t\in\mathbb{T}}, \mathbf{P})$ *be a filtered product economy and assume* $\mathbf{P} = p\,\mu$*. Then we have* $\mathbf{P}$*-almost everywhere*

$$\mathrm{E}(g \mid \mathscr{F}_t)_{|\omega_{\mathbf{B}}^t} = \frac{\int_{\Omega_{\mathbf{F}}^t} g(\omega)p(\omega)\,\mathrm{d}\mu_{\mathbf{F}}^t}{\int_{\Omega_{\mathbf{F}}^t} p(\omega)\,\mathrm{d}\mu_{\mathbf{F}}^t}.$$

Proof By Definition 2.11 the conditional expectation satisfies

$$\int_{\Omega} g(\omega)Z\big(\omega_{\mathbf{B}}^t\big)p(\omega)\,\mathrm{d}\mu = \int_{\Omega} \mathrm{E}(g \mid \mathscr{F}_t)\big(\omega_{\mathbf{B}}^t\big)Z\big(\omega_{\mathbf{B}}^t\big)p(\omega)\,\mathrm{d}\mu$$

for any integrable $\mathscr{F}_t$-measurable function $\omega_{\mathbf{B}}^t \mapsto Z(\omega_{\mathbf{B}}^t)$, and is uniquely determined by this condition. The functions

$$\tilde{g}\big(\omega_{\mathbf{B}}^t\big) = \int_{\Omega_{\mathbf{F}}^t} g(\omega)p(\omega)\,\mathrm{d}\mu_{\mathbf{F}}^t$$

$$\tilde{p}\big(\omega_{\mathbf{B}}^t\big) = \int_{\Omega_{\mathbf{F}}^t} p(\omega)\,\mathrm{d}\mu_{\mathbf{F}}^t$$

are clearly $\mathscr{F}_t$-measurable. Thus their quotient is also $\mathscr{F}_t$-measurable. We make the calculation

$$\begin{aligned}\int_{\Omega} g(\omega)Z\big(\omega_{\mathbf{B}}^t\big)p(\omega)\,\mathrm{d}\mu &= \int_{\Omega_{\mathbf{B}}^t} \tilde{g}\big(\omega_{\mathbf{B}}^t\big)Z\big(\omega_{\mathbf{B}}^t\big)\,\mathrm{d}\mu_{\mathbf{B}}^t \\ &= \int_{\Omega_{\mathbf{B}}^t} \frac{\tilde{g}(\omega_{\mathbf{B}}^t)}{\tilde{p}(\omega_{\mathbf{B}}^t)}\tilde{p}\big(\omega_{\mathbf{B}}^t\big)\,Z\big(\omega_{\mathbf{B}}^t\big)\,\mathrm{d}\mu_{\mathbf{B}}^t\end{aligned}$$

$$= \int_{\Omega} \frac{\tilde{g}(\omega_{\mathbf{B}}^{t})}{\tilde{p}(\omega_{\mathbf{B}}^{t})} Z\big(\omega_{\mathbf{B}}^{t}\big) p(\omega)\, \mathrm{d}\mu.$$

The assertion thus follows immediately from the definition of conditional expectation. □

2.4.2 General Dynamic Risk Measures

Definition 2.16 Let $(\Omega, \mathbf{P})$ be a probability space and $(\mathscr{F}_t)_{t\in\{0,\dots,n\}}$ a filtration on Ω. For each t, let $\mathscr{M}_t(\Omega, \mathbb{R})$ be a vector space of $\mathscr{F}_t$-measurable functions. A *dynamic risk measure* is a family $(\rho_t)_{t\in\{0,\dots,n\}}$ of maps

$$\rho_t : \mathscr{M}_n(\Omega, \mathbb{R}) \to \mathscr{M}_t(\Omega, \mathbb{R}),$$

so that we have:

(i) $X_1 \geq X_2$ a.s. $\Rightarrow \rho_t(X_1) \geq \rho_t(X_2)$ a.s. (Monotony);
(ii) For $K \in \mathscr{M}_t(\Omega, \mathbb{R})$ there holds $\rho_t(X + K) = \rho_t(X) + K$ a.s. (Translation invariance).

Definition 2.17 A dynamic risk measure $(\rho_t)_{t\in\{0,\dots,n\}}$ is called *coherent*, if:

(i) $\rho_t(KX) = K\rho_t(X)$ a.s. for all $K \in \mathscr{M}_t(\Omega, \mathbb{R})$ with $K \geq 0$ a.s., and $KX \in \mathscr{M}_n(\Omega, \mathbb{R})$ (Homogeneity);
(ii) $\rho_t(X_1 + X_2) \leq \rho_t(X_1) + \rho_t(X_2)$ a.s. (Subadditivity).

The definitions are analogous to Definition 2.8. The map ρ_t is the risk measure determined at time t, which has the time horizon $]t, n[$. Since ρ_t is based on the evolution of risk up to the time t, ρ_t is not real-valued, but rather an $\mathscr{M}_t(\Omega, \mathbb{R})$-valued map. On the same grounds, K is taken as an element of $\mathscr{M}_t(\Omega, \mathbb{R})$.

Remark 2.10 In the literature it is often taken that $\mathscr{M}_t(\Omega, \mathbb{R}) = L^\infty(\Omega, \mathscr{F}_t)$ (see e.g., [8]).

In Sect. 2.4.3, we will examine dynamic risk measures on filtered product economies, and thus explicitly employ the product structure. This will lead us to dynamic risk measures with properties relevant to actual practice. In Sect. 2.4.4 we describe an alternative approach for general filtrations, that is favored in the mathematical literature. We shall nonetheless see that this alternative approach is problematic from a practical point of view. Section 2.4.4 can therefore be skipped by readers who are primarily interested in applications.

2.4.3 Dynamic Risk Measures on Filtered Product Economies

Let $(\Omega, (\mathscr{F}_t)_{t\in\mathbb{T}}, \mathbf{P})$ be a filtered product economy and $\rho\colon \mathscr{M}(\Omega, \mathbb{R}) \to \mathbb{R}$ a risk measure. If ρ is "sufficiently generic" then it can be transferred "pointwise", in a natural way, to the fibers $F_t(\omega) \subseteq \Omega$, where the probability measure $\mathbf{P}_{\omega^t_{\mathbf{B}}}$ is used instead of $\mathbf{P}$. This produces, for each fiber $F_t(\omega)$, a risk measure $\rho_t(\omega)\colon \mathscr{M}(F_t(\omega), \mathbb{R}) \to \mathbb{R}$. Since the t-fiber through $\tilde{\omega} \in F_t(\omega)$ is just $F_t(\tilde{\omega}) = F_t(\omega)$, one would also expect that $\rho_t(\omega) = \rho_t(\tilde{\omega})$. The conjecture suggests itself that $(\omega, X) \mapsto \rho_t(\omega)(X(\omega^t_{\mathbf{B}}, \cdot))$ defines a dynamic risk measure. In Theorem 2.6 this general construction will be carried out for the value at risk, and in Theorem 2.7 for the expected shortfall.

Definition 2.18 Let $(\Omega, (\mathscr{F}_t)_{t\in\mathbb{T}}, \mathbf{P})$ be a filtered product economy, $\alpha \in]0, 1[$, and $\mathscr{M}_t(\Omega, \mathbb{R})$ the space of $\mathscr{F}_t$-measurable functions. The family of maps parametrized by $t \in \mathbb{T}$

$$\mathrm{VaR}_{\alpha,t}\colon \mathscr{M}_n(\Omega, \mathbb{R}) \to \mathscr{M}_t(\Omega, \mathbb{R}), \quad X \mapsto \mathrm{VaR}_{\alpha,t}(X)$$

with

$$\mathrm{VaR}_{\alpha,t}(X)_{|\omega^t_{\mathbf{B}}} = \inf\left\{x \in \mathbb{R}\colon \mathbf{P}_{\omega^t_{\mathbf{B}}}\left(X\left(\omega^t_{\mathbf{B}}, \cdot\right) \le x\right) \ge \alpha\right\}$$

is called *dynamic value at risk*.

Theorem 2.6 *The dynamic value at risk is a dynamic risk measure.*

Proof For each $\omega \in \Omega$, $\mathrm{VaR}_{\alpha,t}$ is just the ordinary value at risk for the probability space $(F_t(\omega), \mathbf{P}_{\omega^t_{\mathbf{B}}})$. Because the inequality $X > Y$ carries over in a trivial way to subsets, the monotone property of the value at risk carries over pointwise to $\mathrm{VaR}_{\alpha,t}$ for each $\omega^t_{\mathbf{B}} \in \Omega^t_{\mathbf{B}}$.

Since $\mathscr{F}_t$-measurable functions are constant on the sets

$$\left\{\omega^t_{\mathbf{B}}\right\} \times \Omega^t_{\mathbf{F}} \subseteq \Omega,$$

the same pointwise argument also gives translation invariance.

It remains to show the $\mathscr{F}_t$-measurability of $\mathrm{VaR}_{\alpha,t}(X)$. First we assume that X is bounded below. Then there is an increasing sequence $\{X_k\}_{k\in\mathbb{N}}$ of simple functions with $\limsup_{k\to\infty} X_k = \lim_{k\to\infty} X_k = X$. Since X_k is measurable and takes on only finitely many values, the map $\omega^t_{\mathbf{B}} \mapsto \mathrm{VaR}_{\alpha,t}(X_k)_{|\omega^t_{\mathbf{B}}}$ is likewise measurable. By the monotony of $\mathrm{VaR}_{\alpha,t}$, $\{\mathrm{VaR}_{\alpha,t}(X_k)\}_{k\in\mathbb{N}}$ is an increasing sequence of measurable simple functions, which is why

$$\lim_{k\to\infty} \mathrm{VaR}_{\alpha,t}(X_k)$$

is measurable. On the grounds that $X_k \le X$ and of monotony, we have

$$\lim_{k\to\infty} \mathrm{VaR}_{\alpha,t}(X_k) \le \mathrm{VaR}_{\alpha,t}(X).$$

Assume there is an $\varepsilon > 0$ with $\lim_{k\to\infty} \mathrm{VaR}_{\alpha,t}(X_k) < \mathrm{VaR}_{\alpha,t}(X) - \varepsilon$. Since X_k is an increasing sequence, for all k it is true that $\mathrm{VaR}_{\alpha,t}(X_k) < \mathrm{VaR}_{\alpha,t}(X) - \varepsilon$. This implies

$$\mathbf{P}_{\omega^t_{\mathbf{B}}}\big(X_k\big(\omega^t_{\mathbf{B}},\,\cdot\big) \le \mathrm{VaR}_{\alpha,t}(X) - \varepsilon/2\big) \ge \alpha$$

for all k. From X_k's convergence to X follows

$$\mathbf{P}_{\omega^t_{\mathbf{B}}}\big(X\big(\omega^t_{\mathbf{B}},\,\cdot\big) \le \mathrm{VaR}_{\alpha,t}(X) - \varepsilon/2\big) \ge \alpha$$

contradicting the definition of $\mathrm{VaR}_{\alpha,t}$.

If X is not bounded below, let $\tilde{X}_k = \max(X, -k)$. Then $\{\hat{X}_k\}_{k\in\mathbb{N}}$ is a decreasing sequence of measurable functions bounded below that converges pointwise to X. Therefore $\lim_{k\to\infty} \mathrm{VaR}_{\alpha,t}(\tilde{X}_k)$ is measurable, and we have

$$\lim_{k\to\infty} \mathrm{VaR}_{\alpha,t}(\tilde{X}_k) \ge \mathrm{VaR}_{\alpha,t}(X).$$

If there were an $\varepsilon > 0$ with $\liminf_{k\to\infty} \mathrm{VaR}_{\alpha,t}(\tilde{X}_k) > \mathrm{VaR}_{\alpha,t}(X) + \varepsilon$, then we would have

$$\mathbf{P}_{\omega^t_{\mathbf{B}}}\big(\tilde{X}_k\big(\omega^t_{\mathbf{B}},\,\cdot\big) \le \mathrm{VaR}_{\alpha,t}(X)_{|\omega^t_{\mathbf{B}}} + \varepsilon/2\big) < \alpha$$

for all k, and by the convergence of $\tilde{X}_k$

$$\mathbf{P}_{\omega^t_{\mathbf{B}}}\big(X\big(\omega^t_{\mathbf{B}},\,\cdot\big) \le \mathrm{VaR}_{\alpha,t}(X)_{|\omega^t_{\mathbf{B}}} + \varepsilon/2\big) < \alpha$$

in contradiction to the definition of $\mathrm{VaR}_{\alpha,t}$. □

Definition 2.19 Let $(\Omega, (\mathscr{F}_t)_{t\in\mathbb{T}}, \mathbf{P})$ be a filtered product economy, $\alpha \in]0,1[$ and $\mathscr{M}_t(\Omega,\mathbb{R})$ the space of integrable, $\mathscr{F}_t$-measurable functions. The family of maps parametrized by $t \in \mathbb{T}$

$$\mathrm{ES}_{\alpha,t} : \mathscr{M}_n(\Omega,\mathbb{R}) \to \mathscr{M}_t(F_t,\mathbb{R}), \quad X \mapsto \mathrm{ES}_{\alpha,t}(X)$$

with

$$\mathrm{ES}_{\alpha,t}(X)_{|\omega^t_{\mathbf{B}}} = \frac{1}{1-\alpha}\,\mathrm{E}_{\mathbf{P}_{\omega^t_{\mathbf{B}}}}\left(X\big(\omega^t_{\mathbf{B}},\,\cdot\big)\,1_{X(\omega^t_{\mathbf{B}},\,\cdot),\mathrm{VaR}_{\alpha,t}(X)_{|\omega^t_{\mathbf{B}}},\alpha}\right)$$

is called *dynamic expected shortfall*, where $1_{X,x,\alpha}$ is defined as in Lemma 2.5.

Theorem 2.7 *The dynamic expected shortfall is a coherent, dynamic risk measure.*

Proof For each $\omega \in \Omega$, $\mathrm{ES}_{\alpha,t}$ is the ordinary expected shortfall on the probability space $(F_t(\omega), \mathbf{P}_{\omega^t_{\mathbf{B}}})$. Therefore monotony and subadditivity carry over directly to $\mathrm{ES}_{\alpha,t}(X)$. Since $\mathscr{F}_t$-measurable functions on the sets

$$\big\{\omega^t_{\mathbf{B}}\big\} \times \Omega^t_{\mathbf{F}} \subseteq \Omega$$

are constant, the same pointwise argument delivers translation invariance and homogeneity.

To show the $\mathscr{F}_t$-measurability of $\mathrm{ES}_{\alpha,t}(X)$, we first assume that X is bounded below. Then there is an increasing sequence $\{X_k\}_{k\in\mathbb{N}}$ of simple functions for which $\lim_{k\to\infty} X_k = X$ almost everywhere. Since X_k and $\mathrm{VaR}_{\alpha,t}(X_k)$ are measurable and take on only finitely many values, for each $w \in \pi_t(\omega)$ the map

$$\omega_{\mathbf{B}}^t \mapsto 1_{X_k(\omega_{\mathbf{B}}^t, w), \mathrm{VaR}_{\alpha,t}(X_k), \alpha},$$

and thus also

$$\omega_{\mathbf{B}}^t \mapsto \mathrm{ES}_{\alpha,t}(X_k)_{|\omega_{\mathbf{B}}^t},$$

are $\mathscr{F}_t$-measurable. Due to the monotony of $\mathrm{ES}_{\alpha,t}$, $\{\mathrm{ES}_{\alpha,t}(X_k)\}_{k\in\mathbb{N}}$ is an increasing sequence of measurable, simple functions, from which we see

$$\limsup_{k\to\infty} \mathrm{ES}_{\alpha,t}(X_k) = \limsup_{k\to\infty} \frac{\int_{F_t(\omega_{\mathbf{B}}^t)} 1_{X_k(\omega_{\mathbf{B}}^t, \cdot), \mathrm{VaR}_{\alpha,t}(X_k), \alpha}\, p(\omega_{\mathbf{B}}^t, \cdot)\, \mathrm{d}\mu_{\mathbf{F}}^t}{(1-\alpha)\int_{F_t(\omega_{\mathbf{B}}^t)} p(\omega_{\mathbf{B}}^t, \cdot)\, \mathrm{d}\mu_{\mathbf{F}}^t} = \mathrm{ES}_{\alpha,t}(X)$$

is measurable.

If X is not bounded from below, we can consider the sequence

$$\tilde{X}_k = \big\{\max(X, -k)\big\}_{k\in\mathbb{N}}.$$

We have just seen that $\mathrm{ES}_{\alpha,t}(\tilde{X}_k)$ is measurable for each k. Therefore also measurable is the function

$$\begin{aligned}\liminf_{k\to\infty} \mathrm{ES}_{\alpha,t}(\tilde{X}_k) &= \liminf_{k\to\infty} \frac{\int_{F_t(\omega_{\mathbf{B}}^t)} 1_{X_k(\omega_{\mathbf{B}}^t, \cdot), \mathrm{VaR}_{\alpha,t}(\max(X,-k)), \alpha}\, p(\omega_{\mathbf{B}}^t, \cdot)\, \mathrm{d}\mu_{\mathbf{F}}^t}{(1-\alpha)\int_{F_t(\omega_{\mathbf{B}}^t)} p(\omega_{\mathbf{B}}^t, \cdot)\, \mathrm{d}\mu_{\mathbf{F}}^t} \\ &= \mathrm{ES}_{\alpha,t}(X).\end{aligned}$$

□

It is conceivable, that a badly chosen risk measure can, over time, lead to contradictory estimates of risk. In the following definition we therefore formalize a minimum requirement upon dynamic risk measures, that should not be violated by risk measures used in practice.

Definition 2.20 Let (Ω, μ) be a measure space, $B \subset \Omega$ and $f : \Omega \to \bar{\mathbb{R}}$ a map. The *essential supremum* of f over B is defined by

$$\operatorname{ess\,sup}_B(f) = \inf\big\{a \in \mathbb{R} : \mu\big(\{x : f(x) > a\} \cap B\big) = 0\big\} \in \bar{\mathbb{R}}$$

and the *essential infimum* of f over B by

$$\operatorname{ess\,inf}_B(f) = \sup\big\{a \in \mathbb{R} : \mu\big(\{x : f(x) < a\} \cap B\big) = 0\big\} \in \bar{\mathbb{R}}.$$

Definition 2.21 Let $(\mathscr{F}_t)_{t\in\mathbb{T}}$ be a product filtration on $\Omega = \prod_{s=1}^n \Omega_s$. A dynamic risk measure $(\rho_t)_{t\in\mathbb{T}}$ is *time-consistent*, if for each random variable X, almost every $\omega \in \Omega$ and each $\mathscr{F}_{t+1}$-measurable subset $B \subseteq F_t(\omega)$ with $\mathbf{P}_{\omega_{\mathbf{B}}^t}(B) > 0$ and

$$\operatorname{ess\,inf}_B\big(\rho_{t+1}(X)\big) > \rho_t(X)_{|\omega},$$

there is an $\mathscr{F}_{t+1}$ -measurable subset $C \subseteq F_t(\omega)$ with $\mathbf{P}_{\omega_{\mathbf{B}}^t}(C) > 0$ and

$$\operatorname{ess\,sup}_C\big(\rho_{t+1}(X)\big) < \rho_t(X)_{|\omega}.$$

For an illustration of Definition 2.21 see Fig. 2.8.

Definition 2.21 expresses the idea that the capital requirement cannot increase in time for each possible future. Let ρ_t be a time-consistent, dynamic risk measure and $\rho_t(X)_{|\omega} = K$. If at time t the probability that $\rho_{t+1}(X)_{|\omega} > K$ is greater than 0, so more capital must be provided, then the probability that $\rho_{t+1}(X)_{|\omega} < K$, so some capital can be freed, is also greater than 0.

Definition 2.22 Let $(\mathscr{F}_t)_{t\in\mathbb{T}}$ be a product filtration on $\Omega = \prod_{s=1}^n \Omega_s$. A dynamic risk measure ρ_t is *weakly time-consistent*, if for almost every ω and each pair (X, t) there holds

$$\operatorname{ess\,inf}_{F_t(\omega)}\big(\rho_{t+1}(X)\big) \le \rho_t(X)_{|\omega}.$$

Remark 2.11 In practice it is impossible to distinguish between time-consistency and weak time-consistency. This is because if ρ_t is weakly time-consistent and X is a given random variable, then an arbitrarily small change in X is enough to make X time-consistent.

Corollary 2.3 *Time-consistency implies weak time-consistency.*

Make the assumption that an enterprise uses, in setting capital levels, a dynamic risk measure ρ_t that violates weak time-consistency at some time τ. Suppose further that at time τ the business puts up sufficient capital $\rho_\tau(X)$. With probability 1 it will have to provide more capital one period later, although in the interim no cash has flowed. Worse risk management is hardly imaginable. On these grounds one will certainly want to require of any risk measure to be used in practice that it is time-consistent.

Remark 2.12 In the literature the notion "time-consistent" is often used, to express another condition in which two random variables are compared with each other (see Definition 2.24). We will, however, see in Sect. 2.4.4, that the condition in Definition 2.24 is less useful for practical risk management than requiring time-consistency.

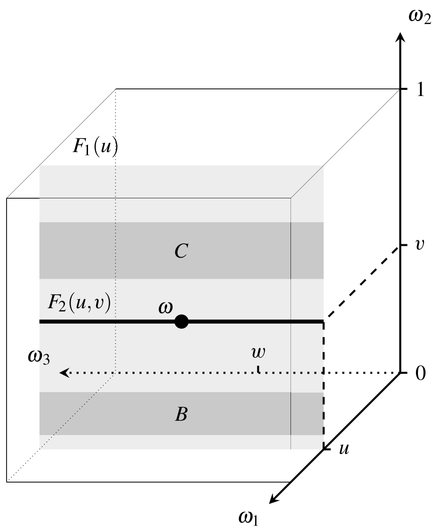

Fig. 2.8 Illustration of Definition 2.21. We consider the space $\Omega =]0, 1[^3$, where $\mathbf{P}$ is the Lebesgue measure, and write $\omega = (u, v, w)$. The quantity $\rho_1(u)$ is constant on $F_1(u)$ and lies between the values of the risk measure ρ_2 on B and C

Remark 2.13 Remark that Definition 2.21 and Definition 2.22 are not time-symmetric. There are good reasons for this, since the risk should lessen as the time horizon shrinks, as in less time fewer losses can occur.

Example 2.9 Let $\Omega =]0, 1[\times]0, 1[$, and $\mathbf{P}$ be Lebesgue measure on Ω and $\alpha \in]0, 1[$. Assume the random variable X is defined by

$$X(\omega) = \begin{cases} 2 & \text{if } \omega_1 < (1-\alpha)/2, \\ 1 & \text{otherwise.} \end{cases}$$

Then we have $\mathrm{VaR}_{\alpha,0}(X)_{|\omega} = 1$ for all $\omega \in \Omega$ but

$$\mathrm{VaR}_{\alpha,1}(X)_{|\tilde{\omega}} = \begin{cases} 2 & \text{if } \tilde{\omega}_1 < (1-\alpha)/2, \\ 1 & \text{otherwise.} \end{cases}$$

This shows the dynamic value at risk is not time-consistent.

Theorem 2.8 *Let $(\Omega, (\mathscr{F}_t)_{t\in\mathbb{T}}, \mathbf{P})$ be a filtered product economy. Then the dynamic value at risk on $(\Omega, (\mathscr{F}_t)_{t\in\mathbb{T}}, \mathbf{P})$ is weakly time-consistent.*

Proof Let $\omega \in \Omega$. For $t \in \{0, \dots, n-1\}$ we put

$$G = \left\{\tilde{\omega} \in F_t(\omega) \colon X(\tilde{\omega}) \geq \mathrm{VaR}_{\alpha,t+1}(X)_{|\tilde{\omega}_{\mathbf{B}}^{t+1}}\right\}.$$

Let $\tilde{\omega} \in F_t(\omega)$. Lemma 2.1 implies

$$\mathbf{P}_{\tilde{\omega}_{\mathbf{B}}^{t+1}}\left(X\left(\tilde{\omega}_{\mathbf{B}}^{t+1}, \cdot\right) \leq \mathrm{VaR}_{\alpha,t+1}(X)_{|\tilde{\omega}_{\mathbf{B}}^{t+1}}\right) = \alpha,$$

since $\mathrm{VaR}_{\alpha,t+1}(X)_{|\tilde{\omega}_{\mathbf{B}}^{t+1}}$ is exactly the value at risk for the random variable $X(\tilde{\omega}_{\mathbf{B}}^{t+1},\cdot)$ on the probability space $(F_{t+1}(\tilde{\omega}),\mathbf{P}_{\tilde{\omega}_{\mathbf{B}}^{t+1}})$. It follows that

$$\begin{aligned}1-\alpha &\le \mathbf{P}_{\tilde{\omega}_{\mathbf{B}}^{t+1}}\big(X\big(\tilde{\omega}_{\mathbf{B}}^{t+1},\cdot\big)\ge \mathrm{VaR}_{\alpha,t+1}(X)\big)\\ &= \frac{\int_{\Omega_{\mathbf{F}}^{t+1}} 1_G\, p(\tilde{\omega}_{\mathbf{B}}^{t+1},\cdot)\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}}{\int_{\Omega_{\mathbf{F}}^{t+1}} p(\tilde{\omega}_{\mathbf{B}}^{t+1},\cdot)\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}}\end{aligned}$$

for each $\tilde{\omega}\in F_t(\omega)$. Thus we obtain

$$\begin{aligned}\mathbf{P}_{\omega_{\mathbf{B}}^t}(G)\int_{F_t(\omega)} p\big(\omega_{\mathbf{B}}^t,\cdot\big)\,\mathrm{d}\mu_{\mathbf{F}}^t &= \int_{F_t(\omega)} 1_G\, p\big(\omega_{\mathbf{B}}^t,\cdot\big)\,\mathrm{d}\mu_{\mathbf{F}}^t\\ &= \int_{\Omega_{t+1}}\int_{\Omega_{\mathbf{F}}^{t+1}} 1_G\, p\big(\omega_{\mathbf{B}}^t,\cdot\big)\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}\,\mathrm{d}\mu_{t+1}\\ &\ge \int_{\Omega_{t+1}}(1-\alpha)\int_{\Omega_{\mathbf{F}}^{t+1}} p\big(\omega_{\mathbf{B}}^t,\cdot\big)\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}\,\mathrm{d}\mu_{t+1}\\ &= (1-\alpha)\int_{F_t(\omega)} p\big(\omega_{\mathbf{B}}^t,\cdot\big)\,\mathrm{d}\mu_{\mathbf{F}}^t.\end{aligned}$$

Therefore it holds that $\mathbf{P}_{\omega_{\mathbf{B}}^t}(G)\ge 1-\alpha$ and $X(\tilde{\omega})\ge \operatorname{ess\,inf}_{F_{t+1}(\tilde{\omega})}(\mathrm{VaR}_{\alpha,t+1}(X))$ for almost every $\tilde{\omega}\in G$. This implies $\mathrm{VaR}_{\alpha,t}(X)_{|\omega}\ge \operatorname{ess\,inf}_{F_t(\omega)}(\mathrm{VaR}_{\alpha,t+1}(X))$. □

Theorem 2.9 *Let $(\Omega,(\mathscr{F}_t)_{t\in\mathbb{T}},\mathbf{P})$ be a filtered product economy, then the dynamic expected shortfall on $(\Omega,(\mathscr{F}_t)_{t\in\mathbb{T}},\mathbf{P})$ is time-consistent.*

Proof Assume $\omega\in\Omega$ and $u=\pi_t(\omega)$. We suppose that there are a random variable X and an $\mathscr{F}_{t+1}$-measurable subset $B\subseteq F_t(\omega)$ with $\mathbf{P}_u(B)>0$ and

$$\operatorname{ess\,inf}_B\big(\mathrm{ES}_{\alpha,t+1}(X)\big)>\mathrm{ES}_{\alpha,t}(X)_{|\omega}.$$

If there is no $\mathscr{F}_{t+1}$-measurable set $C\subseteq F_t(\omega)$ with $\mathbf{P}_u(C)>0$ and

$$\operatorname{ess\,sup}_C\big(\mathrm{ES}_{\alpha,t+1}(X)\big)<\mathrm{ES}_{\alpha,t}(X)_{|\omega},$$

then for almost every $v\in\Omega_{t+1}$

$$\mathrm{ES}_{\alpha,t+1}(X)_{|(u,v)}\ge \mathrm{ES}_{\alpha,t}(X)_{|u}. \tag{2.7}$$

It suffices therefore to show that the inequality (2.7) leads to a contradiction.

We assume that (2.7) holds and put

$$\begin{aligned}G&=\big\{\tilde{\omega}\in F_t(\omega):\ X(\tilde{\omega})>\mathrm{VaR}_{\alpha,t+1}(X)_{|\pi_{t+1}(\tilde{\omega})}\big\},\\ H&=\big\{\tilde{\omega}\in F_t(\omega):\ X(\tilde{\omega})=\mathrm{VaR}_{\alpha,t+1}(X)_{|\pi_{t+1}(\tilde{\omega})}\big\}\end{aligned}$$

as well as

$$\beta\colon \Omega_{t+1} \to [0,1],$$
$$v \mapsto \beta(v) = \begin{cases} \frac{1-\alpha-\mathbf{P}_{(u,v)}(G\cap F_{t+1}(\omega))}{\mathbf{P}_{(u,v)}(H\cap F_{t+1}(\omega))} & \text{if } \mathbf{P}_{(u,v)}(H\cap F_{t+1}(\omega)) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that because of the $(t+1)$-fiberwise definition of G and H we have in general $G \neq \{X > \mathrm{VaR}_{\alpha,t}(X)\}$ and $H \neq \{X = \mathrm{VaR}_{\alpha,t}(X)\}$.

By Lemma 2.5, there holds for all $v \in \Omega_{t+1}$

$$\begin{aligned}
&(1-\alpha)\ \mathrm{ES}_{\alpha,t+1}(X)_{|(u,v)} \\
&= \int_{F_{t+1}(u,v)} X\left(1_G + \beta(v)\,1_H\right) \mathrm{d}\mathbf{P}_{(u,v)} \\
&= \frac{\int_{F_{t+1}(u,v)} X\,1_G\,p\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}}{\int_{F_{t+1}(u,v)} p\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}} + \frac{\int_{F_{t+1}(u,v)} X\,\beta(v)\,1_H\,p\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}}{\int_{F_{t+1}(u,v)} p\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}}.
\end{aligned}$$

Since the inequality (2.7) has been assumed, on the set $B \subset F_t(\omega)$ strict inequality actually holds, and B is not a $\mathbf{P}_u$-null set, we obtain by integration over v the inequality

$$\begin{aligned}
&(1-\alpha)\ \mathrm{ES}_{\alpha,t}(X)_{|u} \int_{F_t(u)} p\,\mathrm{d}\mu_{\mathbf{F}}^{t} \\
&< \int_{F_t(u)} X\,1_G\,p\,\mathrm{d}\mu_{\mathbf{F}}^{t} \\
&\quad + \int_{\Omega_{t+1}} \beta \int_{F_{t+1}(u,\cdot)} X\,1_H\,p\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}\,\mathrm{d}\mu_{t+1} \\
&= \left(\int_G X\,\mathrm{d}\mathbf{P}_u + \int_H \beta X\,\mathrm{d}\mathbf{P}_u\right) \int_{F_t(u)} p\,\mathrm{d}\mu_{\mathbf{F}}^{t}. \qquad (2.8)
\end{aligned}$$

From

$$\beta(v) \int_{F_{t+1}(u,v)} 1_H\,p\,\mathrm{d}\mu_{\mathbf{F}}^{t+1} = (1-\alpha)\int_{F_{t+1}(u,v)} p\,\mathrm{d}\mu_{\mathbf{F}}^{t+1} - \int_{F_{t+1}(u,v)} 1_G\,p\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}$$

follows

$$\begin{aligned}
\int_{\Omega_{t+1}} \beta \int_{F_{t+1}(u,\cdot)} 1_H\,p\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}\,\mathrm{d}\mu_{t+1} &= (1-\alpha)\int_{\Omega_{t+1}} \int_{F_{t+1}(u,\cdot)} p\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}\,\mathrm{d}\mu_{t+1} \\
&\quad - \int_{\Omega_{t+1}} \int_{F_{t+1}(u,\cdot)} 1_G\,p\,\mathrm{d}\mu_{\mathbf{F}}^{t+1}\,\mathrm{d}\mu_{t+1} \\
&= (1-\alpha)\int_{F_t(u)} p\,\mathrm{d}\mu_{\mathbf{F}}^{t} - \int_{F_t(u)} 1_G\,p\,\mathrm{d}\mu_{\mathbf{F}}^{t}
\end{aligned}$$

and from this $\int_G \mathrm{d}\mathbf{P}_u + \int_H \beta\,\mathrm{d}\mathbf{P}_u = 1-\alpha$. Let

$$V = \left\{\tilde\omega \in F_t(\omega) \colon\ X(\tilde\omega) > \mathrm{VaR}_{\alpha,t}(X)_{|u}\right\},$$
$$c = \begin{cases} 1 - \frac{\int_{H\cap V}(1-\beta)\,\mathrm{d}\mathbf{P}_u}{\int_{H\setminus V}\beta\,\mathrm{d}\mathbf{P}_u} & \text{if } \int_{H\setminus V}\beta\,\mathrm{d}\mathbf{P}_u > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} 1-\alpha &= \int_G \mathrm{d}\mathbf{P}_u + \int_H \beta\,\mathrm{d}\mathbf{P}_u \\ &= \int_G \mathrm{d}\mathbf{P}_u + c\int_{H\setminus V}\beta\,\mathrm{d}\mathbf{P}_u + \int_{H\cap V}\mathrm{d}\mathbf{P}_u \\ &= \int_{(G\cup H)\cap V}\mathrm{d}\mathbf{P}_u + \int_{G\setminus V}\mathrm{d}\mathbf{P}_u + c\int_{H\setminus V}\beta\,\mathrm{d}\mathbf{P}_u. \end{aligned} \tag{2.9}$$

We now show that, in addition,

$$\int_H \beta X\,\mathrm{d}\mathbf{P}_u \le c\int_{H\setminus V}\beta X\,\mathrm{d}\mathbf{P}_u + \int_{H\cap V} X\,\mathrm{d}\mathbf{P}_u. \tag{2.10}$$

From $\inf_{\tilde\omega\in H\cap V} X(\tilde\omega) \ge \sup_{\tilde\omega\in H\setminus V} X(\tilde\omega)$ and $0\le\beta\le 1$ follows

$$\begin{aligned} \int_{H\cap V} X\,\mathrm{d}\mathbf{P}_u &\ge \int_{H\cap V}\Big(\inf_{\tilde\omega\in H\cap V} X(\tilde\omega)(1-\beta) + \beta X\Big)\,\mathrm{d}\mathbf{P}_u \\ &\ge \sup_{\tilde\omega\in H\setminus V} X(\tilde\omega)\left(\int_{H\cap V}(1-\beta)\,\mathrm{d}\mathbf{P}_u\right) + \int_{H\cap V}\beta X\,\mathrm{d}\mathbf{P}_u \end{aligned}$$

and thus

$$\begin{aligned} \int_{H\cap V}\beta X\,\mathrm{d}\mathbf{P}_u \int_{H\setminus V}\beta\,\mathrm{d}\mathbf{P}_u &\le \int_{H\cap V} X\,\mathrm{d}\mathbf{P}_u \int_{H\setminus V}\beta\,\mathrm{d}\mathbf{P}_u \\ &\quad - \sup_{\tilde\omega\in H\setminus V} X(\tilde\omega)\left(\int_{H\cap V}(1-\beta)\,\mathrm{d}\mathbf{P}_u\right)\int_{H\setminus V}\beta\,\mathrm{d}\mathbf{P}_u \\ &\le \int_{H\cap V} X\,\mathrm{d}\mathbf{P}_u \int_{H\setminus V}\beta\,\mathrm{d}\mathbf{P}_u \\ &\quad - \int_{H\setminus V}\beta X\,\mathrm{d}\mathbf{P}_u \int_{H\cap V}(1-\beta)\,\mathrm{d}\mathbf{P}_u. \end{aligned}$$

We conclude

$$\begin{aligned}\int_H \beta X \,\mathrm{d}\mathbf{P}_u \int_{H\setminus V} \beta \,\mathrm{d}\mathbf{P}_u &= \int_{H\setminus V} \beta X \,\mathrm{d}\mathbf{P}_u \int_{H\setminus V} \beta \,\mathrm{d}\mathbf{P}_u + \int_{H\cap V} \beta X \,\mathrm{d}\mathbf{P}_u \int_{H\setminus V} \beta \,\mathrm{d}\mathbf{P}_u \\ &\le \int_{H\setminus V} \beta X \,\mathrm{d}\mathbf{P}_u \Big(\int_{H\setminus V} \beta \,\mathrm{d}\mathbf{P}_u - \Big(\int_{H\cap V} (1-\beta) \,\mathrm{d}\mathbf{P}_u \Big)\Big) \\ &\quad + \int_{H\cap V} X, \mathrm{d}\mathbf{P}_u \int_{H\setminus V} \beta \,\mathrm{d}\mathbf{P}_u \\ &= \int_{H\setminus V} \beta \,\mathrm{d}\mathbf{P}_u \left(c \int_{H\setminus V} \beta X \,\mathrm{d}\mathbf{P}_u + \int_{H\cap V} X \,\mathrm{d}\mathbf{P}_u \right),\end{aligned}$$

which implies the inequality (2.10).

Since H and G are disjoint, it follows from the inequalities (2.8) and (2.10) that

$$\begin{aligned}(1-\alpha)\ \mathrm{ES}_{\alpha,t}(X)_{|u} &< \int_{G\cup(H\cap V)} X \,\mathrm{d}\mathbf{P}_u + c \int_{H\setminus V} \beta\, X \,\mathrm{d}\mathbf{P}_u \\ &= \int_{(G\cup H)\cap V} X \,\mathrm{d}\mathbf{P}_u + \int_{G\setminus V} X \,\mathrm{d}\mathbf{P}_u + c \int_{H\setminus V} \beta\, X \,\mathrm{d}\mathbf{P}_u \\ &\le \int_{(G\cup H)\cap V} X \,\mathrm{d}\mathbf{P}_u + \inf_{\tilde\omega\in V} X(\tilde\omega) \Big(\int_{G\setminus V} \mathrm{d}\mathbf{P}_u + c \int_{H\setminus V} \beta \,\mathrm{d}\mathbf{P}_u \Big) \\ &\overset{(*)}{=} \int_{(G\cup H)\cap V} X \,\mathrm{d}\mathbf{P}_u + \inf_{\tilde\omega\in V} X(\tilde\omega) \Big(1-\alpha - \int_{(G\cup H)\cap V} \mathrm{d}\mathbf{P}_u \Big) \\ &\overset{(**)}{\le} \int_V X \,\mathrm{d}\mathbf{P}_u + \mathrm{VaR}_{\alpha,t}(X)_{|u} \Big(1-\alpha - \int_V \mathrm{d}\mathbf{P}_u \Big) \\ &\overset{(**)}{=} (1-\alpha)\ \mathrm{ES}_{\alpha,t}(X)_{|u},\end{aligned}$$

where we have additionally used in $(*)$ the equation (2.9) and in $(**)$ the definition of the set V. This is a contradiction, so that our assumption that

$$\mathrm{ES}_{\alpha,t+1}(X)_{|(u,v)} \ge \mathrm{ES}_{\alpha,t}(X)_{|u} \quad \text{for almost all } v \in \Omega_{t+1}$$

must have been false. □

This demonstrates that both the dynamic value at risk and the dynamic expected shortfall are good candidates for multi-period risk management.

2.4.4 A Class of Dynamic Risk Measures on General Filtrations

In the pertinent literature, another class of dynamic risk measures is studied that are defined in an elegant way on general filtrations. Proposition 2.9 gives the method of constructing this class. Example 2.10 provides, in our view, a reason why this construction, in spite of its elegance, is hardly suitable for practical risk management.

For Proposition 2.9 we need a modification, for probability spaces, of the (pointwise) supremum over a set of functions.

Definition 2.23 Let $(\Omega, \mathscr{A}, \mu)$ be a measure space, and $\mathscr{S} \subset \mathscr{B}(\Omega, \mathbb{R})$ a subset of measurable functions. Then the *essential supremum*

$$\operatorname{ess\,sup}(\mathscr{S}) \in \mathscr{B}(\Omega, \bar{\mathbb{R}})$$

is defined by the following properties:

(i) $\operatorname{ess\,sup}(\mathscr{S}) \geq f$ holds a.s. for any $f \in \mathscr{S}$.
(ii) If $g \in \mathscr{B}(\Omega, \mathbb{R})$ satisfies the inequality $g \geq f$ a.s for each $f \in \mathscr{S}$, then it follows that $g \geq \operatorname{ess\,sup}(\mathscr{S})$ a.s.

Remark that $\operatorname{ess\,sup}(\mathscr{S})$ can take the value ∞ on a set with genuinely positive measure.

Remark 2.14 Observe that the equation $\operatorname{ess\,sup}\{f\} = f$ holds for any f. Hence for singleton sets of functions Definition 2.23 does not reduce to Definition 2.20.

Lemma 2.14 *Let* $(\Omega, \mathscr{A}, \mathbf{P})$ *be a probability space and* $\mathscr{S} \subset \mathscr{B}(\Omega, \mathbb{R})$, $\mathscr{S} \neq \emptyset$. *Then there exists* $\operatorname{ess\,sup}(\mathscr{S}) \in \mathscr{B}(\Omega, \bar{\mathbb{R}})$.

Proof Let

$$\tilde{\mathscr{S}} = \left\{ \omega \mapsto \max_{g \in S} \{g(\omega)\} : S \subseteq \mathscr{S} \text{ is a finite set} \right\}$$

be the set of pointwise maxima of all finite subsets of $\mathscr{S}$ and

$$\theta(x) = \begin{cases} -\frac{\pi}{2} & \text{when } x = -\infty, \\ \arctan(x) & \text{when } -\infty < x < \infty, \\ \frac{\pi}{2} & \text{when } x = \infty. \end{cases}$$

Since $\mathbf{P}$ is a probability measure, and for any measurable function g the composition $\theta \circ g$ is bounded and measurable, there exists $\int_\Omega |\theta \circ g| \, \mathrm{d}\mathbf{P}$. From this we see $\theta \circ g$ is integrable, and

$$\alpha = \sup \left\{ \int_\Omega \theta \circ g \, \mathrm{d}\mathbf{P} : \, g \in \tilde{\mathscr{S}} \right\} \in \bar{\mathbb{R}}$$

is well-defined. Let $\{g_k\}_{k \in \mathbb{N}} \subseteq \tilde{\mathscr{S}}$ be a sequence with

$$\lim_{k \to \infty} \int_\Omega \theta \circ g_k \, \mathrm{d}\mathbf{P} = \alpha.$$

Since for any k we can replace the function g_k with $\max\{g_1, \dots, g_k\}$, we can assume without any loss of generality that $g_{k+1} \geq g_k$ for each $k \in \mathbb{N}$. Since the sequence $\{g_k\}_{k\in\mathbb{N}}$ is increasing, we have for $f(\omega) = \sup_{k\in\mathbb{N}} g_k(\omega)$ and for each $\omega \in \Omega$

$$f(\omega) = \lim_{k\in\mathbb{N}} g_k(\omega) \in \bar{\mathbb{R}}.$$

f is also measurable as the supremum of measurable functions, because of which $\int_\Omega \theta \circ f \,\mathrm{d}\mathbf{P}$ is well defined. Since θ is bounded, it follows from the theorem of dominated convergence and the continuity of θ that

$$\alpha = \lim_{k\to\infty} \int_\Omega \theta \circ g_k \,\mathrm{d}\mathbf{P} = \int_\Omega \lim_{k\to\infty} \theta \circ g_k \,\mathrm{d}\mathbf{P} = \int_\Omega \theta \circ \lim_{k\to\infty} g_k \,\mathrm{d}\mathbf{P} = \int_\Omega \theta \circ f \,\mathrm{d}\mathbf{P}.$$

Let $g \in \mathscr{S}$. Since $\max(g_k, g) \in \tilde{\mathscr{S}}$ for any $k \in \mathbb{N}$ and

$$\max\{f, g\} = \max\left\{\lim_{k\to\infty} g_k, g\right\} = \lim_{k\to\infty} \max\{g_k, g\},$$

it follows that

$$\int_\Omega \theta \circ \max\{f, g\} \,\mathrm{d}\mathbf{P} = \int_\Omega \lim_{k\to\infty} \theta \circ \max\{g_k, g\} \,\mathrm{d}\mathbf{P} = \lim_{k\to\infty} \int_\Omega \theta \circ \max\{g_k, g\} \,\mathrm{d}\mathbf{P} \leq \alpha.$$

Thus the integral over $\theta \circ f - \theta \circ \max\{f, g\}$ is non-negative. However,

$$\theta \circ \max\{f, g\} - \theta \circ f$$

is also non-negative since θ is monotone. This is only possible if $\theta \circ \max\{f, g\} = \theta \circ f$ a.s. It follows that $f \geq g$ a.s.

Now let $g \in \mathscr{B}(\Omega, \mathbb{R})$ be a measurable function that satisfies the inequality $g \geq h$ a.s for any $h \in \mathscr{S}$. For each $k \in \mathbb{N}$ then it obviously holds that $g \geq g_k$. It follows that $g \geq \sup_{k\in\mathbb{N}} g_k = f$, and so $f = \operatorname{ess\,sup}(\mathscr{S})$. □

The expected shortfall of a random variable X can be expressed as the supremum of the expectations of X relative to a class of probability measures (Proposition 2.3). The following proposition shows that this sort of representation leads in a natural way to coherent risk measures.

Proposition 2.8 *Let $\mathscr{W}$ be a subset of the probability measures on a probability space $(\Omega, \mathscr{A}, \mathbf{P})$ with the following properties*:

(i) *For each $X \in \mathscr{M}(\Omega, \mathbb{R})$ and each $Q \in \mathscr{W}$ there exists* $\mathrm{E}_Q(X)$.
(ii) $\mathbf{Q} \ll \mathbf{P}$ *for all* $\mathbf{Q} \in \mathscr{W}$.

Then a coherent risk measure is defined by

$$\rho^{\mathscr{W}}(X) = \sup_{\mathbf{Q}\in\mathscr{W}} \left\{\mathrm{E}_{\mathbf{Q}}(X)\right\}.$$

We will prove Proposition 2.8 after Proposition 2.9 as an elementary corollary.

If one chooses $\mathscr{M}_n(\Omega, \mathbb{R})$ to be the space of almost everywhere measurable functions, then one can show that a large class of coherent risk measures can be represented according to Proposition 2.8 [4, Theorem 3.2]. This motivates also constructing coherent dynamic risk measures using this representation.

Proposition 2.9 *Let $(\Omega, \mathscr{A}, \mathbf{P})$ be a probability space and $(\mathscr{F}_t)_{t\in\{0,\dots,n\}}$ a filtration with $\mathscr{F}_n = \mathscr{A}$. For $t \in \{0, \dots, n\}$ let $\mathscr{M}_t(\Omega, \mathbb{R})$ be the vector space of a.e. bounded, $\mathscr{F}_t$-measurable functions. Let $\mathscr{W}$ be a set of probability measures on Ω.*

Then the family of maps

$$\rho_t^{\mathscr{W}} : \mathscr{M}_n(\Omega, \mathbb{R}) \to \mathscr{M}_t(\Omega, \mathbb{R}), \quad X \mapsto \rho_t^{\mathscr{W}}(X) = \operatorname{ess} \sup_{\mathbf{Q}\in\mathscr{W}} \mathrm{E}_{\mathbf{Q}}(X \mid \mathscr{F}_t)$$

is a coherent, dynamic risk measure.

Proof We show first that $\rho_t^{\mathscr{W}}$ is a dynamic risk measure.

By Lemma 2.14 the map $\rho_t^{\mathscr{W}}(X)$ exists and it is $\mathscr{F}_t$-measurable. Since X is almost everywhere bounded, we have $\rho_t^{\mathscr{W}}(X)_{|\omega} \in \mathbb{R}$ a.s.

Monotony: Suppose $X_1 \geq X_2$ a.s. Then for each $\mathbf{Q} \in \mathscr{W}$ the inequality $\mathrm{E}_{\mathbf{Q}}(X_1) \geq \mathrm{E}_{\mathbf{Q}}(X_2)$ holds. Since the relationship "$\geq$" is preserved in passing to the supremum the monotone property of $\rho_t^{\mathscr{W}}$ follows.

Translation invariance: Let $K \in \mathscr{M}_t(\Omega, \mathbb{R})$ and $\mathbf{Q} \in \mathscr{W}$. Since K is $\mathscr{F}_t$-measurable, we have

$$\mathrm{E}_{\mathbf{Q}}(X + K \mid \mathscr{F}_t) = \mathrm{E}_{\mathbf{Q}}(X \mid \mathscr{F}_t) + K.$$

Passing to the supremum then brings $\rho_t^{\mathscr{W}}(X + K) = \rho_t^{\mathscr{W}}(X) + K$.

Homogeneity: Let $K \in \mathscr{M}_t(\Omega, \mathbb{R})$ with $K \geq 0$ a.s. and $KX \in \mathscr{M}_n(\Omega, \mathbb{R})$. For $\mathbf{Q} \in \mathscr{W}$ it then holds that

$$\mathrm{E}_{\mathbf{Q}}(KX \mid \mathscr{F}_t) = K\,\mathrm{E}_{\mathbf{Q}}(X \mid \mathscr{F}_t)$$

and for this that

$$\rho_t^{\mathscr{W}}(KX) = \sup_{\mathbf{Q}\in\mathscr{W}} \left\{K\,\mathrm{E}_{\mathbf{Q}}(X) \mid \mathscr{F}_t\right\} = K \sup_{\mathbf{Q}\in\mathscr{W}} \left\{\mathrm{E}_{\mathbf{Q}}(X \mid \mathscr{F}_t)\right\} = K\rho_t^{\mathscr{W}}(X),$$

where we have used $K \geq 0$ in deriving the second equality.

Subadditivity: For $X_1, X_2 \in \mathscr{M}_n(\Omega, \mathbb{R})$ we have

$$\begin{aligned} \sup_{\mathbf{Q}\in\mathscr{W}} \left\{\mathrm{E}_{\mathbf{Q}}(X_1 + X_2 \mid \mathscr{F}_t)\right\} &= \sup_{\mathbf{Q}\in\mathscr{W}} \left\{\mathrm{E}_{\mathbf{Q}}(X_1 \mid \mathscr{F}_t) + \mathrm{E}_{\mathbf{Q}}(X_2 \mid \mathscr{F}_t)\right\} \\ &\leq \sup_{\mathbf{Q}\in\mathscr{W}} \left\{\mathrm{E}_{\mathbf{Q}}(X_1 \mid \mathscr{F}_t)\right\} + \sup_{\mathbf{Q}\in\mathscr{W}} \left\{\mathrm{E}_{\mathbf{Q}}(X_2 \mid \mathscr{F}_t)\right\}. \end{aligned}$$ □

Proof of Proposition 2.8 We apply Proposition 2.9 for

$$n = 1, \qquad \mathscr{F}_1 = \mathscr{A}, \qquad \mathscr{F}_0 = \{\emptyset, \Omega\}.$$

Then $\rho_0^{\mathscr{W}}(X) = \operatorname{ess\,sup}_{\mathbf{Q}\in\mathscr{W}} \mathrm{E}_{\mathbf{Q}}(X \mid \mathscr{F}_0) = \sup_{\mathbf{Q}\in\mathscr{W}} \mathrm{E}_{\mathbf{Q}}(X) = \rho^{\mathscr{W}}(X)$ and it immediately follows from the definition of a coherent dynamic risk measure that ρ_0 is a coherent risk measure. $\square$

It would seem mathematically plausible, to extend the expected shortfall to a dynamic risk measure using Proposition 2.9 for a given filtration. In fact, this route has been traveled in the literature (see e.g. [8, Example 25]). Unfortunately, this extension is not well suited for practical applications, since the extended risk measure loses the characteristics of the expected shortfall (see Example 2.10).

Example 2.10 We consider the probability space

$$(\Omega, \mathbf{P}) = \big(]0,1[\times]0,1[, \mathrm{d}\omega\big),$$

where Ω is equipped with the Borel algebra $\mathscr{B}(]0,1[\times]0,1[)$ and $\mathbf{P} = \mathrm{d}\omega = \mathrm{d}\omega_1 \otimes \mathrm{d}\omega_2$ is the Lebesgue measure on $\mathbb{R}^2$. We choose the product filtration

$$\mathscr{F}_0 = \{\emptyset, \Omega\}, \qquad \mathscr{F}_1 = \big\{A\times]0,1[\colon A \in \mathscr{B}\big(]0,1[\big)\big\}, \qquad \mathscr{F}_2 = \mathscr{B}\big(]0,1[\times]0,1[\big)$$

For $\alpha \in]0,1[$ let

$$\mathscr{W}_\alpha = \left\{\mathbf{Q} \colon \mathbf{Q} \text{ is a probability measure with } \mathbf{Q} \ll \mathbf{P} \text{ and } \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}} \le \frac{1}{1-\alpha}\right\}.$$

Then by Proposition 2.3 it holds that $\rho_0^{\mathscr{W}_\alpha} = \mathrm{ES}_\alpha$ and $\rho_t^{\mathscr{W}_\alpha}$ is the coherent dynamic risk measure derived from the expected shortfall using Proposition 2.9.

For $\mu \in]0,1[$ let

$$\begin{aligned}
A_\mu &= \left\{\omega \colon 0 < \omega_1 < \frac{1}{4},\ 0 < \omega_2 < 2(1-\mu)(1-\alpha)\right\},\\
B_\mu &= \left\{\omega \colon \frac{3}{4} \le \omega_1 < 1,\ 0 < \omega_2 < 2(1-\mu)(1-\alpha)\right\},\\
C_\mu &= \{\omega \colon 1-\mu+\mu\alpha \le \omega_2 < 1\}
\end{aligned}$$

(see the Fig. 2.9). By construction, we have $A_\mu \cap B_\mu = B_\mu \cap C_\mu = C_\mu \cap A_\mu = \emptyset$ as well as

$$\mathbf{P}(A_\mu) = \mathbf{P}(B_\mu) = \frac{1}{2}(1-\mu)(1-\alpha), \qquad \mathbf{P}(C_\mu) = \mu(1-\alpha),$$

so that $\mathbf{P}(A_\mu \cup B_\mu \cup C_\mu) = 1-\alpha$. Therefore the measure $\mathbf{Q}_\mu$ defined by the density

$$\frac{\mathrm{d}\mathbf{Q}_\mu}{\mathrm{d}\mathbf{P}} = \frac{1_{A_\mu} + 1_{B_\mu} + 1_{C_\mu}}{1-\alpha}$$

must be a probability measure and satisfies the inequality

$$\frac{d\mathbf{Q}_\mu}{d\mathbf{P}} \le \frac{1}{1-\alpha}.$$

We have $\mathbf{Q}_\mu \in \mathscr{W}_\alpha$ and thus for all bounded random variables X the inequality

$$\mathrm{E}_{\mathbf{Q}_\mu}(X \mid \mathscr{F}_t) \le \rho_t^{\mathscr{W}_\alpha}(X).$$

We now look at the random variable

$$X\colon \Omega \to \mathbb{R}, \quad (\omega_1, \omega_2) \mapsto \xi(\omega_2),$$

where ξ depends only on ω_2 and is monotone increasing. For $\omega_1 \in]\frac{1}{4}, \frac{3}{4}[$ it holds by Proposition 2.7 that

$$\begin{aligned}
\mathrm{E}_{\mathbf{Q}_\mu}(X \mid \mathscr{F}_1)_{|\omega_1} &= \frac{\int_0^1 \xi(\omega_2) \frac{d\mathbf{Q}_\mu}{d\mathbf{P}}(\omega_1, \omega_2)\, d\omega_2}{\int_0^1 \frac{d\mathbf{Q}_\mu}{d\mathbf{P}}(\omega_1, \omega_2)\, d\omega_2} \\
&= \frac{\frac{1}{1-\alpha}\int_{1-\mu(1-\alpha)}^1 \xi(\omega_2)\, d\omega_2}{\frac{1}{1-\alpha}\int_{1-\mu(1-\alpha)}^1 d\omega_2} \\
&= \frac{1}{\mu(1-\alpha)} \int_0^{\mu(1-\alpha)} \xi(1-x)\, dx
\end{aligned}$$

and thus

$$\lim_{\mu\to 0} \mathrm{E}_{\mathbf{Q}_\mu}(X \mid \mathscr{F}_1)_{|\omega_1} = \lim_{x\to 0} \xi(1-x) = \sup_{\omega_2 \in]0,1[} \xi(\omega_2) = \sup_{\omega\in\Omega} X(\omega).$$

Now let $\omega_1 \in]0, \frac{1}{4}[\cup]\frac{3}{4}, 1[$ and

$$\tilde{A}_\mu = \frac{1}{4} + A_\mu, \qquad \tilde{B}_\mu = -\frac{1}{4} + B_\mu.$$

We can repeat the same analysis with the risk measure $\tilde{\mathbf{Q}}_\mu$ given by

$$\frac{d\tilde{\mathbf{Q}}_\mu}{d\mathbf{P}} = \frac{1_{\tilde{A}_\mu} + 1_{\tilde{B}_\mu} + 1_{C_\mu}}{1-\alpha}$$

and obtain

$$\lim_{\mu\to 0} \mathrm{E}_{\tilde{\mathbf{Q}}_\mu}(X \mid \mathscr{F}_1)_{|\omega_1} = \sup_{\omega\in\Omega} X(\omega)$$

for all $\omega_1 \in]0, \frac{1}{4}[\cup]\frac{3}{4}, 1[$. Because $\rho_1^{\mathscr{W}_\alpha}(X) \le \sup_{\omega\in\Omega} X(\omega)$ it follows that $\rho_1^{\mathscr{W}_\alpha}(X) = \sup_{\omega\in\Omega} X(\omega)$ almost surely. This means that the dynamic risk measure

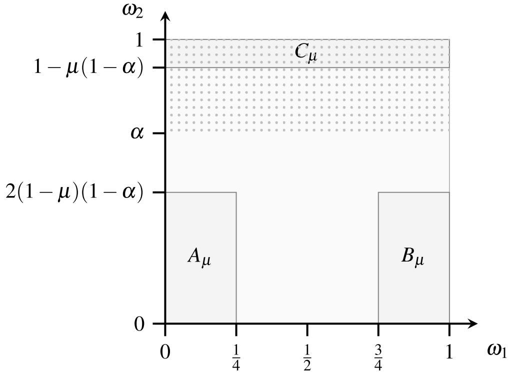

Fig. 2.9 The construction of the measure $\mathbf{Q}_\mu$ in Example 2.10

has completely lost, at time 1, the character of the expected shortfall. This risk measure is naturally totally unsuited for risk management at time $t = 1$. If ξ is not almost certainly constant, then, in addition, there follows the existence of a $c > 0$, so that

$$\rho_0^{\mathscr{W}_\alpha}(X) + c < \rho_1^{\mathscr{W}_\alpha}(X) \quad \text{a.s.}$$

The risk measure $\rho_t^{\mathscr{W}_\alpha}$ is not weakly time-consistent and thus also not time-consistent.

Observe that this example is simple but not at all pathological: the critical feature of our example is that for each $\omega \in \Omega$ we can find a probability measure $\mathbf{Q} \in \mathscr{W}_\alpha$ which, restricted to $F_t(\omega)$, does not vanish only in a small neighborhood of the supremum of X. Since $F_t(\omega)$ is a null set relative to $\mathbf{P}$ we have enough room on $\Omega \setminus F_t(\omega)$ to extend the measure $\mathbf{Q}$ so that we have $\mathbf{Q} \in \mathscr{W}_\alpha$. It follows that $\rho_t^{\mathscr{W}_\alpha}(X) = \sup_{\hat{\omega} \in F_t(\omega)}\{X(\hat{\omega})\}$. This construction can be carried out for nearly any practical example.

In the literature it is often concluded that one must be very cautious in using the expected shortfall in a dynamic context, or that the expected shortfall is simply unsuited to dynamic contexts (see for example [2, Sect. 5.3], [8, Example 25]). The authors limit themselves to a subclass of dynamic risk measures that satisfy additional axiomatically introduced conditions of time-consistence. A typical example of such a condition is the following (see [7, 8]).

Definition 2.24 A dynamic risk measure ρ_t is called *comparison consistent*, if for all random variables $X, Y \in \mathscr{M}(\Omega, \mathbb{R})$ it holds that

$$\rho_{t+1}(X) \geq \rho_{t+1}(Y) \text{ a.s.} \quad \Rightarrow \quad \rho_t(X) \geq \rho_t(Y) \text{ a.s.}$$

Remark 2.15 In [7, 8] the author speaks of "time-consistent" instead of "comparison consistent". We are, however, of the opinion that the consistency condition in Definition 2.24 is not plausible. Since the risk measure at the start of the time period t does not provide a full description of the future cash flow, it could happen that as

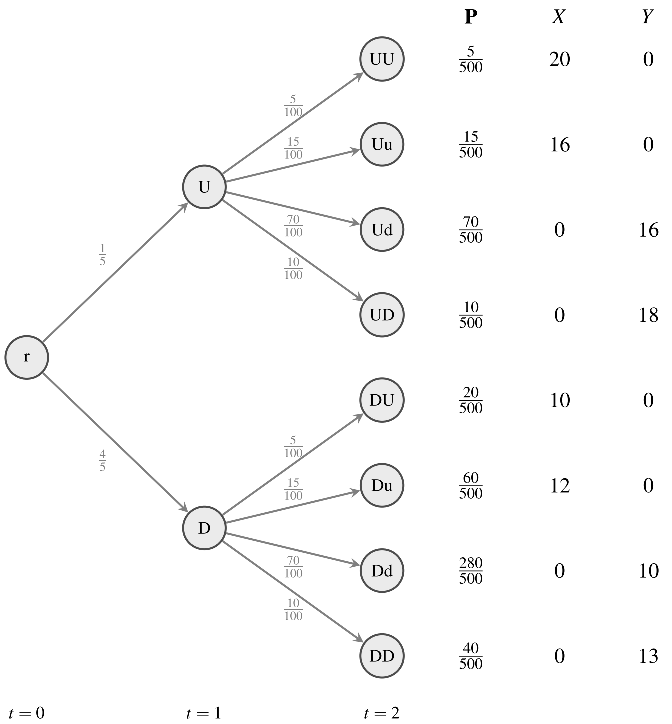

Fig. 2.10 Breakdown of comparison consistency for the expected shortfall

a result of new information at the start of period $t+1$ the risk of the random quantity X is larger than that associated with Y, even if in the earlier time period their relationship was the opposite. Example 2.11 shows how such a gain in information can lead to a violation of comparison consistency.

Therefore we shall reserve "time-consistent" for the variant given in Definition 2.21.

The expected shortfall has properties that have shown themselves useful in the risk management in the single-period case. The extended expected shortfall derived from using Proposition 2.9 fundamentally changes its nature if one breaks up the period into several partial periods. This is not a deficiency of the expected shortfall but the fault of the special construction in the extension. In fact, the generalization given in the Definition 2.19 of the expected shortfall is time-consistent and retains its character with passage of time.

For a dynamic risk measure of the form $\rho_t^{\mathscr{W}}(X) = \operatorname{ess\,sup}_{\mathbf{Q}\in\mathscr{W}} \mathrm{E}_{\mathbf{Q}}(X \mid \mathscr{F}_t)$ the defining set $\mathscr{W}$ does not depend on the time t. It is therefore difficult to describe any gain in information about the risk environment that may happen with time. This seems to be the key problem with this construction and is independent of the choice of the special risk measure "expected shortfall".

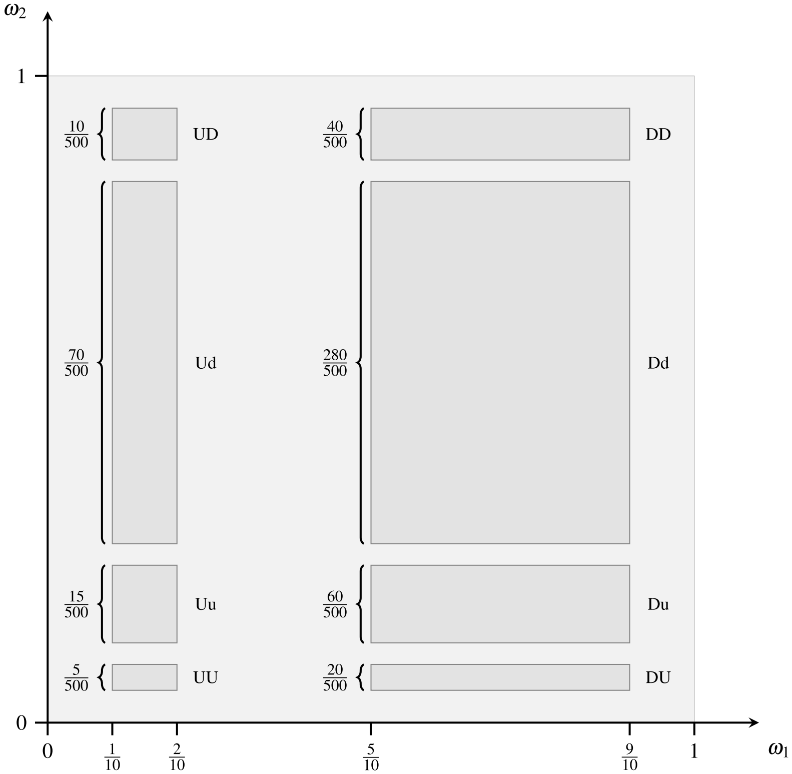

Fig. 2.11 Construction of the discrete Example 2.11 using a product filtration with $\Omega =]0,1[\times]0,1[$ and $\mathbf{P} = \frac{5}{2} \times 1_{\mathrm{UU} \cup \mathrm{Uu} \cup \mathrm{Ud} \cup \mathrm{UD} \cup \mathrm{DU} \cup \mathrm{Du} \cup \mathrm{Dd} \cup \mathrm{DD}}$. The filtration is given by $\mathscr{F}_0 = \{\emptyset, \Omega\}$, $\mathscr{F}_1 = \{A \times]0,1[: A \subseteq [0,1[\text{ is Borel-measurable}\}$ and $\mathscr{F}_2 = \mathscr{B}(]0,1[\times]0,1[)$

The following example shows that the expected shortfall fails to be comparison consistent, even though it is time-consistent.

Example 2.11 Figure 2.10 shows an example in which the comparison consistency of the expected shortfall $\mathrm{ES}_{80\,\%,t}$ breaks down. Note that this discrete example can be constructed using a filtered product economy based on a Borel algebra (see Fig. 2.11). From Lemma 2.5 we obtain at time $t = 1$ the following values:

$$\mathrm{ES}_{80\,\%,1}(X)_{|(U)} = \frac{20 \times 5 + 16 \times 15}{20} = \frac{340}{20} = 17,$$

$$\mathrm{ES}_{80\,\%,1}(X)_{|(D)} = \frac{12 \times 15 + 10 \times 5}{20} = \frac{230}{20} = 11.5,$$

$$\mathrm{ES}_{80\,\%,1}(Y)_{|(U)} = \frac{18 \times 10 + 16 \times 10}{20} = \frac{340}{20} = 17,$$

$$\mathrm{ES}_{80\,\%,1}(Y)_{|(D)} = \frac{13 \times 10 + 10 \times 10}{20} = \frac{230}{20} = 11.5,$$

so that the values of X and Y at time 1 agree. Comparison consistency would then imply that one also has

$$\mathrm{ES}_{80\,\%,0}(X)_{|(r)} = \mathrm{ES}_{80\,\%,0}(Y)_{|(r)}.$$

However, direct calculation for the time $t = 0$ gives

$$\mathrm{ES}_{80\,\%,0}(X)_{|(r)} = \frac{20 \times 5 + 16 \times 15 + 12 \times 60 + 10 \times 20}{100} = \frac{1260}{100} = 12.6$$

$$\mathrm{ES}_{80\,\%,0}(Y)_{|(r)} = \frac{18 \times 10 + 16 \times 70 + 13 \times 20}{100} = \frac{1560}{100} = 15.6$$

If we set $\tilde{Y} = Y - c$ for $c \in]0, 3[$, we obtain the apparently stronger statements,

$$\mathrm{ES}_{80\,\%,0}(\tilde{Y}) > \mathrm{ES}_{80\,\%,0}(X), \quad \text{but } \mathrm{ES}_{80\,\%,1}(\tilde{Y}) < \mathrm{ES}_{80\,\%,1}(X) \text{ a.s.}$$

We interpret this result as saying that portfolio $\tilde{Y}$ wins in capital efficiency over portfolio X at time $t = 1$. Comparison consistency is simply lost because in evaluating $\mathrm{ES}_{80\,\%,0}$ both branches, U and D, contribute to the 20 % of highest losses, while at time 1 the choice is of the 20 % largest losses on each of the branches U resp. D. The relevant information gain in node U is thus that the events that bring losses in D cannot occur again. Analogously one gains the information in node D that the loss-bringing events in U cannot occur. So we do not see a consistency violation in the vernacular sense.

References

1. P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risk. Math. Finance **9**(3), 203–228 (1999) (p. 31)
2. P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, H. Ku, Coherent multiperiod risk adjusted values and Bellman's principle. Ann. Oper. Res. **152**(1), 5–22 (2007) (p. 69)
3. H. Bauer, *Measure and Integration Theory*. De Gruyter Studies in Mathematics (de Gruyter, Berlin, 2001) (p. 48)
4. F. Delbaen, Coherent risk measures on general probability spaces, in *Advances in Finance and Stochastics, Essays in Honour of Dieter Sondermann* (Springer, New York, 2002), pp. 1–37 (p. 66)
5. J. Jacod, P.E. Protter, *Probability Essentials*, 2nd edn. Universitext (Springer, Berlin, 2003) (p. 46)
6. A. McNeil, R. Frey, P. Embrechts, *Quantitative Risk Management. Concepts, Techniques, Tools* (Princeton University Press, Princeton, 2005) (p. 36)
7. F. Riedel, Dynamic coherent risk measures, in *Stochastic Processes and Their Applications*, vol. 112, (2004), pp. 185–200 (p. 69)
8. F. Riedel, Dynamic risk measures. Lecture Notes of a Mini-Course at Paris IX, Dauphine, December 2007 (pp. 54, 67, 69)
9. W.R. van Zwet, A strong law for linear functions of order statistics. Ann. Probab. **8**(5), 986–990 (1980) (p. 40)

http://www.springer.com/978-1-4471-6304-6

Value-Oriented Risk Management of Insurance Companies
Kriele, M.; Wolf, J.
2014, XII, 378 p. 48 illus., Softcover
ISBN: 978-1-4471-6304-6