2 Basics concepts in risk management

2.1 Risk management for a financial firm

2.1.1 Assets, liabilities and the balance sheet

A stylized balance sheet for a bank is:

Assets Investments of the firm		Liabilities Obligations from fundraising	
Cash	£10M	Customer deposits	£80M
(and central bank balance)			
Securities	£50M	Bonds issued	
- bonds, stocks, derivatives		- senior bond issues	£25M
Loans and mortgages	£100M	- subordinated bond issues	£15M
- corporates		Short-term borrowing	£30M
- retail and smaller clients		Reserves (for losses on loans)	£20M
- government			
Other assets	£20M	Debt (sum of above)	£170M
- property			
- investments in companies		Equity	£30M
Short-term lending	£20M		
Total	£200M	Total	£200M

A stylized balance sheet for an insurer is:

Assets		Liabilities	
Investments		Reserves for policies written (technical provisions)	£80M
- bonds	£50M		
- stocks	£5M	Bonds issued	£10M
- property	£5M		
Investments for unit-linked contracts	£30M	Debt (sum of above)	£90M
Other assets	£10M	Equity	£10M
- property			
Total	£100M	Total	£100M

- Balance sheet equation: Assets = Liabilities = Debt + Equity. If equity > 0, the company is *solvent*, otherwise *insolvent*.
- Valuation of the items on the balance sheet is a non-trivial task.
  - *Amortized cost accounting* values a position a *book value* at its inception and this is carried forward/progressively reduced over time.

- *Fair-value accounting* values assets at prices they are sold and liabilities at prices that would have to be paid in the market. This can be challenging for non-traded or illiquid assets or liabilities.

There is a tendency in the industry to move towards fair-value accounting. Market consistent valuation in Solvency II follows similar principles.

2.1.2 Risks faced by a financial firm

- Decrease in the value of the investments on the asset side of the balance sheet (e.g. losses from securities trading or credit risk)
- *Maturity mismatch* (large parts of the assets are relatively illiquid (long-term) whereas large parts of the liabilities are rather short-term obligations. This can lead to a default of a solvent bank or a bank run).
- The prime risk for an insurer is *insolvency* (risk that claims of policy holders cannot be met). On the asset side, risks are similar to those of a bank. On the liability side, the main risk is that reserves are insufficient

to cover future claim payments. Note that the liabilities of a life insurer are of a long-term nature and subject to multiple categories of risk (e.g. interest rate risk, inflation risk and longevity risk).

- So risk is found on both sides of the balance sheet and thus RM should not focus on the asset side alone.

2.1.3 Capital

- There are different notions of capital. One distinguishes:

Equity capital
- Value of assets − debt;
- Measures the firm's value to its shareholders;
- Can be split into *shareholder capital* (initial capital invested in the firm) and *retained earnings* (accumulated earnings not paid out to shareholders).

Regulatory capital
- Capital required according to regulatory rules;

	–	For European insurance firms: MCR + SCR;
	–	A regulatory framework also specifies the capital quality. One distinguishes *Tier 1 capital* (i.e. shareholder capital + retained earnings; can act in full as buffer) and *Tier 2 capital* (includes other positions on the balance sheet).
Economic capital	–	Capital required to control the probability of becoming insolvent (typically over one year);
	–	Internal assessment of risk capital;
	–	Aims at a holistic view (assets and liabilities) and works with fair values of balance sheet items.

- All of these notions refer to items on the liability side that entail no obligations to outside creditors; they can thus serve as buffer against losses.

2.2 Modelling value and value change

2.2.1 Mapping of risks

We now set up a general mathematical model for (changes in) value caused by financial risks. For this we work on a *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ and consider a risk or loss as a *random variable* $X : \Omega \to \mathbb{R}$ (or: L).

- Consider a portfolio of assets and possibly liabilities. The *value* of the portfolio at time t (*today*) is denoted by V_t (a random variable; assumed to be known at t; its df is typically not trivial to determine!).
- We consider a given *time horizon* Δt and assume:

  1) the portfolio composition remains fixed over Δt;

  2) there are no intermediate payments during Δt

  $\Rightarrow$ Fine for small Δt but unlikely to hold for large Δ.

- The *change* in value of the portfolio is then given by

$$\Delta V_{t+1} = V_{t+1} - V_t$$

and we define the (random) *loss* by the sign-adjusted value change

$$L_{t+1} = -\Delta V_{t+1}$$

(as QRM is mainly concerned with losses).

Remark 2.1

1) The distribution of L_{t+1} is called *loss distribution* (df F_L or simply F).
2) Practitioners often consider the *profit-and-loss (P&L) distribution* which is the distribution of $-L_{t+1} = \Delta V_{t+1}$.
3) For longer time intervals, $\Delta V_{t+1} = V_{t+1}/(1+r) - V_t$ (r = risk-free interest rate) would be more appropriate, but we will mostly neglect this issue.

- V_t is typically modelled as a function f of time t and a d-dimensional random vector $\boldsymbol{Z} = (Z_{t,1}, \dots, Z_{t,d})$ of *risk factors* (d typically large), that is,

$$V_t = f(t, \boldsymbol{Z}_t) \quad (\textit{mapping} \text{ of risks})$$

  for some measurable $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$. The choice of f and $\boldsymbol{Z}_t$ is problem-specific (but typically known).

- It is often convenient to work with the *risk-factor changes*

$$\boldsymbol{X}_{t+1} = \boldsymbol{Z}_{t+1} - \boldsymbol{Z}_t.$$

  We can rewrite L_{t+1} in terms of $\boldsymbol{X}_t$ via

$$\begin{aligned} L_{t+1} &= -(V_{t+1} - V_t) = -(f(t+1, \boldsymbol{Z}_{t+1}) - f(t, \boldsymbol{Z}_t)) \\ &= -(f(t+1, \boldsymbol{Z}_t + \boldsymbol{X}_{t+1}) - f(t, \boldsymbol{Z}_t)). \end{aligned}$$

  We see that the loss df is determined by the loss df of $\boldsymbol{X}_{t+1}$.

- If f is differentiable, its first-order (Taylor) approximation is

$$f(t+1, \boldsymbol{Z}_t + \boldsymbol{X}_{t+1}) \approx f(t, \boldsymbol{Z}_t) + f_t(t, \boldsymbol{Z}_t) \cdot 1 + \sum_{j=1}^{d} f_{z_j}(t, \boldsymbol{Z}_t) \cdot X_{t+1,j}$$

We can thus approximate L_{t+1} by the *linearized loss*

$$L^{\Delta}_{t+1} = -\Big(\underbrace{f_t(t, \boldsymbol{Z}_t)}_{=: c_t} + \sum_{j=1}^{d} \underbrace{f_{z_j}(t, \boldsymbol{Z}_t)}_{=: b_{t,j}} X_{t+1,j}\Big) = -(c_t + \boldsymbol{b}_t' \boldsymbol{X}_{t+1}),$$

a linear function of $X_{t+1,1}, \dots, X_{t+1,d}$ (indices denote partial derivatives). The approximation is best if the risk-factor changes are small in absolute value.

Example 2.2 (Stock portfolio)

Consider a portfolio $\mathcal{P}$ of d stocks $S_{t,1},\dots,S_{t,d}$ ($S_{t,j}$ = value of stock j at time t) and denote by λ_j the number of shares of stock j in $\mathcal{P}$. In finance and risk management, one typically uses logarithmic prices as risk factors, i.e. $Z_{t,j} = \log S_{t,j}$, $j \in \{1,\dots,d\}$. Then

$$V_t = f(t, \boldsymbol{Z}_t) = \sum_{j=1}^d \lambda_j S_{t,j} = \sum_{j=1}^d \lambda_j e^{Z_{t,j}}.$$

- The one-period ahead loss is then given by

$$\begin{aligned} L_{t+1} &= -(V_{t+1} - V_t) = -\sum_{j=1}^d \lambda_j (e^{Z_{t,j}+X_{t+1,j}} - e^{Z_{t,j}}) \\ &= -\sum_{j=1}^d \lambda_j e^{Z_{t,j}} (e^{X_{t+1,j}} - 1) = -\sum_{j=1}^d \underbrace{\lambda_j S_{t,j}}_{=:\, w_{t,j}} (e^{X_{t+1,j}} - 1) \qquad (1) \end{aligned}$$

  which is non-linear in $X_{t+1,j}$.

- With $f_{z_j}(t, \boldsymbol{Z}_t) = \lambda_j e^{Z_{t,j}} = \lambda_j S_{t,j} = w_{t,j}$, the linearized loss is

$$\begin{aligned} L^{\Delta}_{t+1} &= -\Big(f_t(t, \boldsymbol{Z}_t) + \sum_{j=1}^{d} f_{z_j}(t, \boldsymbol{Z}_t) X_{t+1,j}\Big) = -\Big(0 + \sum_{j=1}^{d} w_{t,j} X_{t+1,j}\Big) \\ &= -\boldsymbol{w}_t' \boldsymbol{X}_{t+1}. \end{aligned}$$

- Note that

$$L^{\Delta}_{t+1} = -(c_t + \boldsymbol{b}_t' \boldsymbol{X}_{t+1})$$

  for $c_t = 0$ and $\boldsymbol{b}_t = \boldsymbol{w}_t$.

- If $\boldsymbol{\mu} = \mathbb{E}\boldsymbol{X}_{t+1}$ and $\Sigma = \operatorname{cov} \boldsymbol{X}_{t+1}$ are known, then expectation and variance of the (linearized) one-period ahead loss are

$$\mathbb{E} L^{\Delta}_{t+1} = -\sum_{j=1}^{d} w_{t,j} \mathbb{E}(X_{t+1,j}) = -\boldsymbol{w}_t' \boldsymbol{\mu},$$

$$\operatorname{var} L^{\Delta}_{t+1} = \operatorname{var}(\boldsymbol{w}_t' \boldsymbol{X}_{t+1}) = \boldsymbol{w}_t' \operatorname{cov}(\boldsymbol{X}_{t+1}) \boldsymbol{w}_t = \boldsymbol{w}_t' \Sigma \boldsymbol{w}_t.$$

Example 2.3 (European call option)

Consider a portfolio consisting of a European call option on a non-dividend-paying stock S_t with maturity T and strike (exercise price) K. The Black–Scholes formula says that today's value is

$$V_t = C^{\mathsf{BS}}(t, S_t; r, \sigma, K, T) = S_t \Phi(d_1) - K e^{-r(T-t)} \Phi(d_2), \tag{2}$$

where

- t is the time in years;
- Φ is the df of $\mathrm{N}(0,1)$;
- r is the continuously compounded risk-free interest rate;
- $d_1 = \frac{\log(S_t/K)+(r+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$ and $d_2 = d_1 - \sigma\sqrt{T-t}$; and
- σ is the annualized volatility of S_t (standard deviation).

While (2) assumes r, σ to be constant, this is often not true in real markets. Hence, besides $\log S_t$, we consider r_t, σ_t as risk factors, so

$$\boldsymbol{Z}_t = (\log S_t, r_t, \sigma_t) \ \Rightarrow \ \boldsymbol{X}_{t+1} = (\log(S_{t+1}/S_t),\ r_{t+1} - r_t,\ \sigma_{t+1} - \sigma_t).$$

This implies that the mapping f (in terms of the risk factors) is given by

$$V_t = C^{\mathsf{BS}}(t, e^{Z_{t,1}}; Z_{t,2}, Z_{t,3}, K, T) =: f(t, \boldsymbol{Z}_t)$$

and the linearized one-day ahead loss (omitting the arguments of C^{BS}) is

$$\begin{aligned} L^{\Delta}_{t+1} &= -\Big(f_t(t, \boldsymbol{Z}_t) + \sum_{j=1}^{3} f_{z_j}(t, \boldsymbol{Z}_t) X_{t+1,j}\Big) \\ &= -(C^{\mathsf{BS}}_t \Delta t + C^{\mathsf{BS}}_{S_t} S_t X_{t+1,1} + C^{\mathsf{BS}}_{r_t} X_{t+1,2} + C^{\mathsf{BS}}_{\sigma_t} X_{t+1,3}). \end{aligned}$$

If our risk management horizon is 1 d (as opposed to 1 y), we need to introduce $\Delta t = 1/250$ here. Note that the *"Greeks"* enter (C^{BS}_t is the *theta* of the option; $C^{\mathsf{BS}}_{S_t}$ the *delta*; $C^{\mathsf{BS}}_{r_t}$ the *rho*; $C^{\mathsf{BS}}_{\sigma_t}$ the *vega*).

For portfolios of derivatives, L^{Δ}_{t+1} can be a rather poor approximation to L_{t+1} $\Rightarrow$ higher-order (Taylor) approximations such as the *delta-gamma-approximation* (second-order) can be used.

2.2.2 Valuation methods

Fair value accounting

The *fair value* of an asset/liability is an estimate of the price which would be received/paid on an active market. One distinguishes:

Level 1 *Mark-to-market*. The fair value of an investment is determined from quoted prices for the same instrument; see Example 2.2.

Level 2 *Mark-to-model with objective inputs*. The fair value of an instrument is determined using quoted prices in active markets for similar instruments or by using valuation techniques/models with inputs based on observable market data; see Example 2.3.

Level 3 *Mark-to-model with subjective inputs*. The fair value of an instrument is determined using valuation techniques/models for which some inputs are not observable in the market (e.g. determining default risk of portfolios of loans to companies for which no CDS spreads are available).

Risk-neutral valuation

- ... is widely used for pricing financial products, e.g. derivatives
- *value* of a financial instrument today = expected discounted values of future cash flows; the expectation is taken w.r.t. the *risk-neutral pricing measure* Q (also called *equivalent martingale measure (EMM)*; it turns discounted prices into martingales, so fair bets) as opposed to the real world/physical measure $\mathbb{P}$.
- An risk-neutral pricing measure is a probability measure Q such that the expectation of the discounted payoff w.r.t. Q equals V_0 (fair bet).
- Risk-neutral valuation at t of a claim H at T is done via the *risk-neutral pricing rule*
$$V_0^H = \mathbb{E}_{Q,t}(e^{-r(T-t)}H), \quad t < T,$$
where $\mathbb{E}_{Q,t}(\cdot)$ denotes expectation w.r.t. Q given the information up to and including time t.
- $\mathbb{P}$ is estimated from historical data; Q is calibrated to market prices.

Example 2.4 (European call option continued)

- Suppose that options with strike K or maturity T are not traded, but other options on the same stock are.
- Under $\mathbb{P}$ the stock price (S_t) is assumed to follow a geometric Brownian motion (GBM) (the so-called *Black–Scholes model*) with dynamics $dS_t = \mu S_t\, dt + \sigma S_t\, dW_t$ for constants $\mu \in \mathbb{R}$ (drift) and $\sigma > 0$ (volatility), and a standard Brownian motion (W_t).
- Under the EMM Q, $(e^{-rt}S_t)$ is a martingale and S_t follows a GBM with drift r and volatility σ.
- The European call option payoff is $H = (S_T - K)^+ = \max\{S_T - K, 0\}$ and the risk-neutral valuation formula may be shown to be

$$V_t = E_t^Q(e^{-r(T-t)}(S_T - K)^+) = C^{\mathrm{BS}}(t, S_t; r, \sigma, K, T), \quad t < T; \tag{3}$$

- One typically uses quoted prices $C^{\mathrm{BS}}(t, S_t; r, \sigma, K^*, T^*)$ (for different K^*, T^*) to infer the unknown σ. Then plug this so-called *implied volatility* into (3).

2.2.3 Loss distributions

Having determined the mapping f (may involve *valuation models*, e.g. Black–Scholes, or numerical approximation), we can identify the following key statistical tasks of QRM:

1) Find a statistical model for $\boldsymbol{X}_{t+1}$ (typically a model for forecasting $\boldsymbol{X}_{t+1}$, estimated based on historical data);
2) Compute/derive the df $F_{L_{t+1}}$ (requires the df of $f(t+1, \boldsymbol{Z}_t + \boldsymbol{X}_{t+1})$);
3) Compute a risk measure from $F_{L_{t+1}}$.

There are three general methods to approach these challenges.

1) Analytical method

Idea: Choose $F_{\boldsymbol{X}_{t+1}}$ and f such that $F_{L_{t+1}}$ can be determined explicitly.

Prime example: *Variance-covariance method*, see RiskMetrics (1996):

Assumption 1 $\boldsymbol{X}_{t+1} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$ (e.g. if $(\boldsymbol{Z}_t)$ is a Brownian motion, $(\boldsymbol{S}_t)$ a geometric Brownian motion)

Assumption 2 $F_{L^{\Delta}_{t+1}}$ is a good approximation to $F_{L_{t+1}}$.

$$L^{\Delta}_{t+1} = -(c_t + \boldsymbol{b}_t' \boldsymbol{X}_{t+1}) \underset{\text{Ass. 1}}{\Rightarrow} L^{\Delta}_{t+1} \sim \mathrm{N}(-c_t - \boldsymbol{b}_t' \boldsymbol{\mu},\ \boldsymbol{b}_t' \Sigma \boldsymbol{b}_t).$$

Advantages:
- $F_{L^{\Delta}_{t+1}}$ explicit ($\Rightarrow$ typically explicit risk measures)
- Easy to implement

Drawbacks: Assumptions. Assumption 1 is unlikely to be realistic for daily (probably also weekly/monthly) data. Stylized facts about $\boldsymbol{X}_{t+1}$ suggest that $F_{\boldsymbol{X}_{t+1}}$ is *leptokurtic* (thinner body, heavier tail than $\mathrm{N}(\boldsymbol{\mu}, \Sigma)$). Thus, $\boldsymbol{X}_{t+1} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$ underestimates the tail of $F_{L_{t+1}}$ and thus risk measures such as VaR.

When dynamic models for $\boldsymbol{X}_{t+1}$ are considered (e.g. time series models), different estimation methods are possible depending on whether we focus on conditional distributions $F_{\boldsymbol{X}_{t+1}|(\boldsymbol{X}_s)_{s\le t}}$ or the equilibrium distribution $F_{\boldsymbol{X}}$ in a stationary model.

2) Historical simulation

Idea: Estimate $F_{L_{t+1}}$ by its *empirical distribution function (edf)*

$$\hat{F}_{L_{t+1},n}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{\tilde{L}_{t-i+1}\le x\}}, \quad x \in \mathbb{R},$$

based on $\tilde{L}_k = -(f(t+1, \boldsymbol{Z}_t + \boldsymbol{X}_k) - f(t, \boldsymbol{Z}_t))$. $\tilde{L}_{t-n+1}, \dots, \tilde{L}_t$ show what would happen to the current portfolio if the past n risk-factor changes were to recur.

Advantages:
- Easy to implement
- No estimation of the distribution of $\boldsymbol{X}_{t+1}$ required

Drawbacks:
- Sufficient data for all risk-factor changes required
- Only considers past losses ("driving a car by looking in the back mirror")

3) Monte Carlo method

Idea: Take any model for $\boldsymbol{X}_{t+1}$, simulate from it, compute the corresponding simulated losses and estimate $F_{L_{t+1}}$ (typically via edf).

Advantages:
- Quite general (applicable to any model of $\boldsymbol{X}_{t+1}$ which is easy to sample)

Drawbacks:
- Unclear how to find an appropriate model for $\boldsymbol{X}_{t+1}$ (any result is only as good as the chosen $F_{\boldsymbol{X}_{t+1}}$)
- Computational cost (every simulation requires to evaluate the portfolio; expensive, e.g. if the latter contains derivatives which are priced via Monte Carlo themselves $\Rightarrow$ Nested Monte Carlo simulations)

So-called *economic scenario generators* (i.e. economically motivated dynamic models for the evolution and interaction of risk factors) used in insurance also fall under the heading of Monte Carlo methods.

2.3 Risk measurement

- A *risk measure* for a financial position with (random) loss L is a real number which measures the "riskiness of L". In the Basel or Solvency context, it is often interpreted as the amount of capital required to make a position with loss L acceptable to an (internal/external) regulator.
- Some reasons for using risk measures in practice:
  - To determine the amount of capital to hold as a buffer against unexpected future losses on a portfolio (in order to satisfy a regulator/manager concerned with the institution's solvency).
  - As a tool for limiting the amount of risk of a business unit (e.g. by requiring that the daily 95% value-at-risk (i.e. the 95%-quantile) of a trader's position should not exceed a given bound).
  - To determine the riskiness (and thus fair premium) of an insurance contract.

2.3.1 Approaches to risk measurement

Existing approaches to measuring risk can be grouped into three categories:

1) **Notional-amount approach**

- oldest approach
- "standardized approaches" of Basel II (e.g. OpRisk) still use it
- *risk of a portfolio* = summed notational values of the securities times their riskiness factor
- Advantages: ▸ simplicity

  Drawbacks: ▸ No differentiation between long and short positions and no netting: the risk of a long position in corporate bonds hedged by an offsetting position in credit default swaps is counted as twice the risk of the unhedged bond position.

- No diversification benefits: risk of a portfolio of loans to many companies = risk of a portfolio where the whole amount is lent to a single company.
- Problems for portfolios of derivatives: notional amount of the underlying can widely differ from the economic value of the derivative position.

2) **Risk measures based on loss distributions**

- Most modern risk measures are characteristics of the underlying (conditional or unconditional) loss distribution over some predetermined time horizon Δt.
- Examples: variance, value-at-risk, expected shortfall (see later)
- Advantages:
  - The concept of a loss distribution makes sense on all levels (from single portfolios to the overall position of a financial institution).

- If estimated properly, loss distributions reflect netting and diversification effects.

Drawbacks:
- Estimates of loss distributions are typically based on past data.
- It is difficult to estimate loss distributions accurately (especially for large portfolios).

  $\Rightarrow$ Risk measures should be complemented by information from scenarios (forward-looking).

3) **Scenario-based risk measures**

- Typically considered in stress testing.
- One considers possible future risk-factor changes (*scenarios*; e.g. a 20% drop in a market index).
- *Risk of a portfolio* = maximum (weighted) loss under all scenarios.

- If $\mathcal{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ denote the risk-factor changes (scenarios) with corresponding weights $\boldsymbol{w} = (w_1, \dots, w_n)$, the risk is

$$\psi_{\mathcal{X},\boldsymbol{w}} = \max_{1 \le i \le n} \{w_i L(\boldsymbol{x}_i)\}, \tag{4}$$

  where $L(\boldsymbol{x})$ denotes the loss the portfolio would suffer if the hypothetical scenario $\boldsymbol{x}$ were to occur. Many risk measures are of the form (4); see *CME SPAN: Standard Portfolio Analysis of Risk* (2010).

- Mathematical interpretation of (4):
  - Assume $L(\mathbf{0}) = 0$ (okay if Δt small) and $w_i \in [0,1]\ \forall\, i$.
  - $w_i L(\boldsymbol{x}_i) = w_i L(\boldsymbol{x}_i) + (1 - w_i) L(\mathbf{0}) = \mathbb{E}_{\mathbb{P}_i}(L(\boldsymbol{X}_i))$ where $\boldsymbol{X}_i \sim \mathbb{P}_i = w_i \delta_{\boldsymbol{x}_i} + (1 - w_i)\delta_{\mathbf{0}}$ ($\delta_{\boldsymbol{x}}$ the Dirac measure at $\boldsymbol{x}$) is a probability measure on $\mathbb{R}^d$.

  Therefore, $\psi_{\mathcal{X},\boldsymbol{w}} = \max\{\mathbb{E}_{\mathbb{P}}(L(\boldsymbol{X})) : \boldsymbol{X} \sim \mathbb{P} \in \{\mathbb{P}_1, \dots, \mathbb{P}_n\}\}$. Such a risk measure is known as a *generalized scenario*; they play an important role in the theory of coherent risk measures.

- Advantages:
  - Useful for portfolios with few risk factors.
  - Useful complementary information to risk measures based on loss distributions (past data).

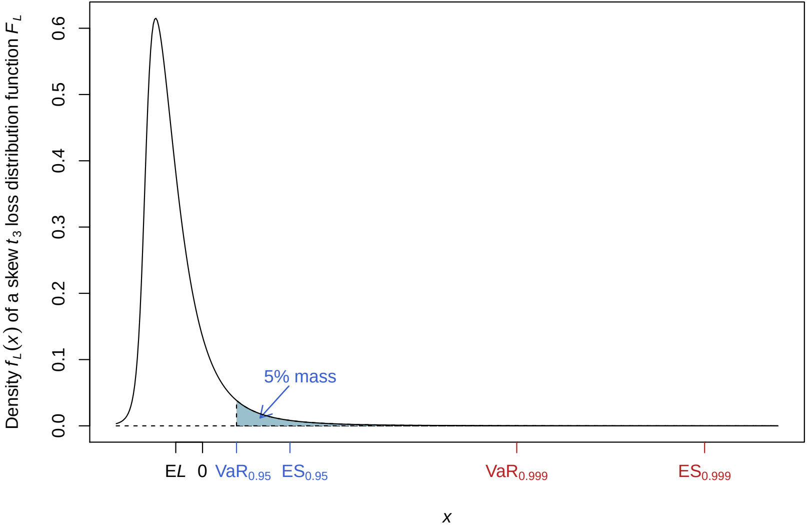
  Drawbacks:
  - Determining scenarios and weights.

2.3.2 Value-at-risk

Definition 2.5 (Value-at-risk)
For a loss $L \sim F_L$, *value-at-risk* (VaR) at confidence level $\alpha \in (0,1)$ is defined by $\mathrm{VaR}_\alpha = \mathrm{VaR}_\alpha(L) = F_L^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F_L(x) \ge \alpha\}$.

- VaR_α is simply the α-quantile of F_L. As such, $F_L(x) < \alpha$ for all $x < \mathrm{VaR}_\alpha(L)$ and $F_L(\mathrm{VaR}_\alpha(L)) = F_L(F_L^{\leftarrow}(\alpha)) \ge \alpha$.
- Known since 1994: Weatherstone 4^{15} report (J.P. Morgan; RiskMetrics)
- VaR is the most widely used risk measure (by Basel II or Solvency II)

- $\mathrm{VaR}_\alpha(L)$ is not a what if risk measure: It does not provide information about the severity of losses which occur with probability $\leq 1-\alpha$



Example 2.6 (VaR_α for $\mathrm{N}(\mu, \sigma^2)$ and $t_\nu(\mu, \sigma^2)$)

1) Let $L \sim \mathrm{N}(\mu, \sigma^2)$. Then

$$F_L(x) = \mathbb{P}(L \le x) = \mathbb{P}((L - \mu)/\sigma \le (x - \mu)/\sigma) = \Phi((x - \mu)/\sigma).$$

This implies that

$$\mathrm{VaR}_\alpha(L) = F_L^{\leftarrow}(\alpha) = F_L^{-1}(\alpha) = \mu + \sigma\Phi^{-1}(\alpha).$$

2) Let $L \sim t_\nu(\mu, \sigma^2)$, so $(L - \mu)/\sigma \sim t_\nu$ and thus, as above,

$$\mathrm{VaR}_\alpha(L) = \mu + \sigma t_\nu^{-1}(\alpha).$$

Note that $X \sim t_\nu = t_\nu(0, 1)$ has density

$$f_X(x) = \frac{\Gamma((\nu + 1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}(1 + x^2/\nu)^{-\frac{\nu+1}{2}}.$$

If $\nu > 1$, $\mathbb{E}X = 0$; if $\nu > 2$, $\operatorname{var} X = \frac{\nu}{\nu-2}$.

Choices of parameters $\Delta t, \alpha$:

- Δt should reflect the time period over which the portfolio is held (unchanged) (e.g. insurance companies: $\Delta t = 1\,\mathsf{y}$)
- Δt should be relatively small (more risk-factor change data is available).
- Typical choices:
  - For limiting traders: $\alpha = 0.95$, $\Delta t = 1\,\mathsf{d}$
  - According to Basel II:
    - Market risk: $\alpha = 0.99$, $\Delta t = 10\,\mathsf{d}$ (2 trading weeks)
    - Credit risk and operational risk: $\alpha = 0.999$, $\Delta t = 1\,\mathsf{y}$
  - According to Solvency II: $\alpha = 0.995$, $\Delta t = 1\,\mathsf{y}$
- Backtesting often needs to be carried out at lower confidence levels in order to have sufficient statistical power to detect poor models.
- Be cautious with strictly interpreting $\mathrm{VaR}_\alpha(L)$ (and other risk measure) estimates, there is typically considerable model/liquidity risk behind.

Interlude: Generalized inverses

$T \nearrow$ means that T is *increasing*, i.e. $T(x) \le T(y)$ for all $x < y$. $T \uparrow$ means that T is *strictly increasing*, i.e. $T(x) < T(y)$ for all $x < y$.

Definition 2.7 (Generalized inverse)

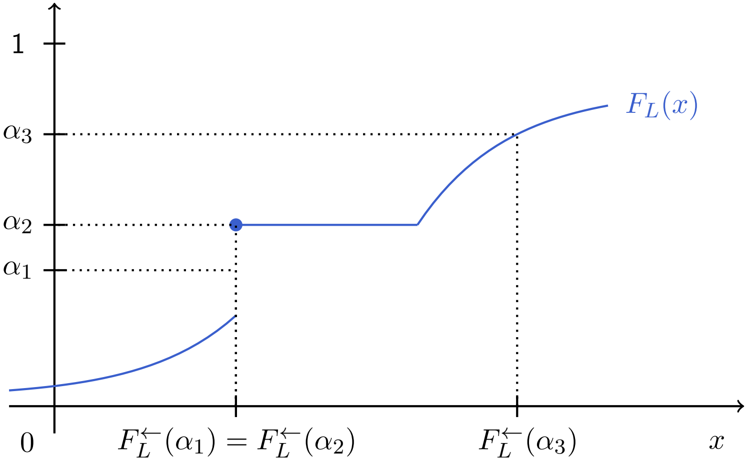
For any increasing function $T : \mathbb{R} \to \mathbb{R}$, with $T(-\infty) = \lim_{x \downarrow -\infty} T(x)$ and $T(\infty) = \lim_{x \uparrow \infty} T(x)$, the *generalized inverse* $T^{\leftarrow} : \mathbb{R} \to \bar{\mathbb{R}} = [-\infty, \infty]$ of T is defined by

$$T^{\leftarrow}(y) = \inf\{x \in \mathbb{R} : T(x) \ge y\}, \quad y \in \mathbb{R},$$

with the convention that $\inf \emptyset = \infty$. If T is a df, $T^{\leftarrow} : [0,1] \to \bar{\mathbb{R}}$ is the *quantile function* of T.

- If T is continuous and $\uparrow$, then $T^{\leftarrow} \equiv T^{-1}$ (ordinary inverse).
- There are rules for working with $T^{\leftarrow}$ (similar to T^{-1}); see Proposition A.15.

$F_L^{\leftarrow}$ visualized for a df F_L:



2.3.3 VaR in risk capital calculations

1) VaR **in regulatory capital calculations for the trading book**

For banks using the *internal model (IM)* approach for market risk in Basel II, the daily risk capital formula is

$$\mathrm{RC}^t = \max\Big\{\mathrm{VaR}_{0.99}^{t,10},\ \frac{k}{60}\sum_{i=1}^{60}\mathrm{VaR}_{0.99}^{t-i+1,10}\Big\} + c.$$

- $\mathrm{VaR}_\alpha^{s,10}$ denotes the 10-day VaR_α calculated at day s ($t =$ today).
- $k \in [3,4]$ is a multiplier (or *stress factor*).
- $c =$ stressed VaR charge (calculated from data from a volatile market period) $+$ incremental risk charge (IRC; $\mathrm{VaR}_{0.999}$-estimate of the annual distribution of losses due to defaults and downgrades) $+$ charges for specific risks.

The averaging tends to lead to smooth changes in the capital charge over time unless $\mathrm{VaR}_{0.99}^{t,10}$ is very large.

2) The Solvency Capital Requirement in Solvency II

The *Solvency Capital Requirement (SCR)* is the amount of capital that enables the insurer to meet its obligations over $\Delta t = 1\,\text{y}$ with $\alpha = 0.995$. Let $V_t = A_t - B_t$ denote the equity capital. The insurer wants to determine the minimum amount of extra capital x_0 to put aside to be solvent in Δt with probability $(\geq)\alpha$. So

$$\begin{aligned}
x_0 &= \inf\{x \in \mathbb{R} : \mathbb{P}(V_{t+1} + x(1+r) \geq 0) \geq \alpha\} \\
&= \inf\Big\{x \in \mathbb{R} : \mathbb{P}\Big(-\Big(\frac{V_{t+1}}{1+r} - V_t\Big) \leq x + V_t\Big) \geq \alpha\Big\} \\
&= \inf\{x \in \mathbb{R} : \mathbb{P}(L_{t+1} \leq x + V_t) \geq \alpha\} \\
&= \inf\{x \in \mathbb{R} : F_{L_{t+1}}(x + V_t) \geq \alpha\} \\
&= \inf\{z - V_t \in \mathbb{R} : F_{L_{t+1}}(z) \geq \alpha\} = \mathrm{VaR}_\alpha(L_{t+1}) - V_t
\end{aligned}$$

and thus $\text{SCR} = V_t + x_0 = \mathrm{VaR}_\alpha(L_{t+1})$ (available capital now + capital required to be solvent in Δt with probability $(\geq)\alpha$). If $x_0 < 0$, the company is already well capitalized.

2.3.4 Other risk measures based on loss distributions

1) Variance (or standard deviation)

- $\mathrm{var}_\alpha(L)$ (or standard deviation) has a long history as a risk measure in finance (due to Markowitz).
- Drawbacks:
  - $\mathbb{E}(L^2) < \infty$ required (not justifiable for non-life insurance or operational risk)
  - no distinction between positive/negative deviations from the mean (var is only a good risk measure if F_L is roughly symmetric around $\mathbb{E}L$, but F_L is typically skewed in credit and operational risk)

2) Expected shortfall

Definition 2.8 (Expected shortfall)
For a loss $L \sim F_L$ with $\mathbb{E}|L| < \infty$, *expected shortfall (*ES*)* at confidence level $\alpha \in (0,1)$ is defined by

$$\mathrm{ES}_\alpha = \mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)\,du. \qquad (5)$$

- ES_α is the average over VaR_u for all $u \ge \alpha$ (if F_L is continuous, ES_α is the average loss beyond VaR_α) $\Rightarrow$ $\mathrm{ES}_\alpha \ge \mathrm{VaR}_\alpha$
- Besides VaR, ES is the most important risk measure in practice.
- ES_α looks further into the tail of F_L, it is a "what if" risk measure (VaR_α is frequency-based; ES_α is severity-based).
- ES_α is more difficult to estimate and backtest than VaR_α (larger sample size required; the variance of estimators is typically larger).
- $\mathrm{ES}_\alpha(L) < \infty$ requires $\mathbb{E}|L| < \infty$.
- If F_L is continuous, $\mathrm{ES}_\alpha(L) = \mathbb{E}(L \,|\, L > \mathrm{VaR}_\alpha(L))$; see the appendix.

- Subadditivity and *elicitability* (see the appendix). One can show:
  - In contrast to VaR_α, ES_α is subadditive (see later).
  - In contrast to ES_α (see Gneiting (2011) or Kou and Peng (2014)), VaR_α exists if $\mathbb{E}|L| = \infty$ and is elicitable (see also the appendix).

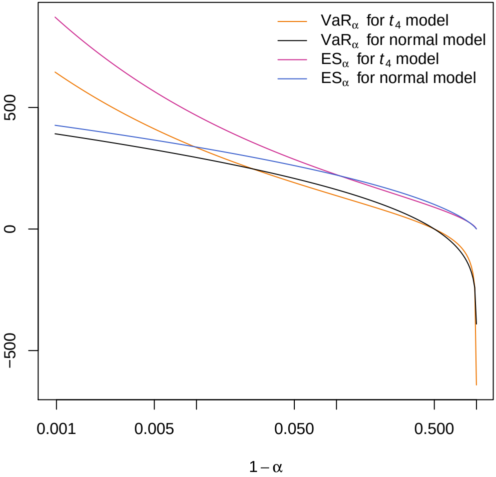
Example 2.9 (VaR and ES for stock returns)

- Consider Example 2.2 with a portfolio consisting of a single stock $V_t = S_t = 10\,000$. In this case, $L^\Delta_{t+1} = -S_t X_{t+1}$, where $X_{t+1} = \log(S_{t+1}/S_t)$.
- Let $\sigma = 0.2/\sqrt{250}$ (annualized volatility of 20%) and assume
  1) $X_{t+1} \sim \mathrm{N}(0, \sigma^2) \Rightarrow L^\Delta_{t+1} \sim \mathrm{N}(0, S_t^2\sigma^2)$;
  2) $X_{t+1} \sim t_\nu(0, \sigma^2\frac{\nu-2}{\nu})$ (so that $\mathrm{var}\, X_{t+1}$ is also σ^2). Then

$$X_{t+1} = \sqrt{\sigma^2\frac{\nu-2}{\nu}}Y \quad \text{for} \quad Y \sim t_\nu.$$

$$\Rightarrow L^\Delta_{t+1} = -S_t\sqrt{\sigma^2\tfrac{\nu-2}{\nu}}Y \sim t_\nu(0, S_t^2\sigma^2\tfrac{\nu-2}{\nu})\ (\mathrm{var}(L^\Delta_{t+1}) = S_t^2\sigma^2).$$

Consider $\nu = 4$ and note that $\mathrm{VaR}_\alpha^{t_4} \geq \mathrm{VaR}_\alpha^{\text{normal}}$ and $\mathrm{ES}_\alpha^{t_4} \geq \mathrm{ES}_\alpha^{\text{normal}}$ only hold for sufficiently large α.



$\Rightarrow$ The t_4 model is not always "riskier" than the normal model.

Example 2.10 (Example 2.6 continued; ES_α for $\mathrm{N}(\mu,\sigma^2)$ and $t_\nu(\mu,\sigma^2)$)

1) Let $\tilde{L} \sim \mathrm{N}(0,1)$. Then $\mathrm{VaR}_\alpha(\tilde{L}) = 0 + 1 \cdot \Phi^{-1}(\alpha)$ and thus

$$\mathrm{ES}_\alpha(\tilde{L}) = \frac{1}{1-\alpha}\int_\alpha^1 \Phi^{-1}(u)\,du \underset{x=\Phi^{-1}(u)}{=} \frac{1}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^\infty x\varphi(x)\,dx,$$

where $\varphi(x) = \Phi'(x) = \exp(-x^2/2)/\sqrt{2\pi}$. Note that $x\varphi(x) = -\varphi'(x)$, so that

$$\mathrm{ES}_\alpha(\tilde{L}) = \frac{-[\varphi(x)]_{\Phi^{-1}(\alpha)}^\infty}{1-\alpha} = \frac{-(0-\varphi(\Phi^{-1}(\alpha)))}{1-\alpha} = \frac{\varphi(\Phi^{-1}(\alpha))}{1-\alpha}.$$

This implies that $L \sim \mathrm{N}(\mu,\sigma^2)$ has expected shortfall

$$\mathrm{ES}_\alpha(L) = \mu + \sigma\,\mathrm{ES}_\alpha(\tilde{L}) = \mu + \sigma\frac{\varphi(\Phi^{-1}(\alpha))}{1-\alpha}.$$

2) Let $L \sim t_\nu(\mu,\sigma^2)$, $\nu > 1$. Similarly as above, one obtains that

$$\mathrm{ES}_\alpha(L) = \mu + \sigma\frac{f_{t_\nu}(t_\nu^{-1}(\alpha))(\nu + t_\nu^{-1}(\alpha)^2)}{(1-\alpha)(\nu-1)},$$

where f_{t_ν} denotes the density of t_ν; see Example 2.6.

By l'Hôpital's Rule (case "0/0"), one can show that

$$1 \le \lim_{\alpha \uparrow 1} \frac{\mathrm{ES}_\alpha(L)}{\mathrm{VaR}_\alpha(L)} = \frac{\nu}{\nu - 1}.$$

- In finance, often $\nu \in (3, 5)$. With $\nu = 3$, $\mathrm{ES}_\alpha(L)$ is 50% larger than $\mathrm{VaR}_\alpha(L)$ (in the limit for large α).
- For $\nu \uparrow \infty$, $\lim_{\alpha \uparrow 1} \frac{\mathrm{ES}_\alpha(L)}{\mathrm{VaR}_\alpha(L)} \downarrow 1$; for $\nu \downarrow 1$, $\lim_{\alpha \uparrow 1} \frac{\mathrm{ES}_\alpha(L)}{\mathrm{VaR}_\alpha(L)} \uparrow \infty$.

Conclusion:

For losses with *heavy tails* (power-like), the difference between using VaR and ES as risk measures for computing risk capital can be huge (for large α as required by Basel II).

2.3.5 Coherent and convex risk measures

- Artzner et al. (1999) (coherent risk measures) and Föllmer and Schied (2002) (convex risk measures) propose axioms of a good risk measure.
- Assume that risk measures ϱ are defined on a linear space of random variables $\mathcal{M}$ (including constants; we can thus add rvs, multiply them with constants etc.), so $\varrho : \mathcal{M} \to \mathbb{R}$.
- There are two possible interpretations of elements of $\mathcal{M}$:

  1) Elements of $\mathcal{M}$ are random variables V_{t+1}; $\tilde{\varrho}(V_{t+1})$ denotes the amount of additional capital that needs to be added to a position with future value V_{t+1} to make it acceptable to a regulator.

  2) Elements of $\mathcal{M}$ are losses $L_{t+1} = -(V_{t+1} - V_t)$; $\varrho(L_{t+1})$ denotes the total amount of capital necessary to back a position with loss L.

  1) and 2) are related via $\varrho(L_{t+1}) = V_t + \tilde{\varrho}(V_{t+1})$ (total capital $=$ available capital $+$ additional capital). We focus on 2) and drop "$t+1$".

Axioms of coherence

Axiom 1 (monotonicity) $L_1, L_2 \in \mathcal{M},\ L_1 \le L_2$ (a.s., i.e. almost surely) $\Rightarrow \varrho(L_1) \le \varrho(L_2)$

Interpr.: Positions which lead to a higher loss in every state of the world require more risk capital.

Criticism: none

Axiom 2 (translation invar.) $\varrho(L + l) = \varrho(L) + l$ for all $L \in \mathcal{M}$, $l \in \mathbb{R}$

Interpr.:

- By adding $l \in \mathbb{R}$ to a position with loss L, we alter the capital requirements accordingly.
- If $\varrho(L) > 0$, and $l = -\varrho(L)$, then $\varrho(L - \varrho(L)) = \varrho(L + l) = \varrho(L) + l = 0$ so that adding $\varrho(L)$ to a position with loss L makes it acceptable.

Criticism: Most people believe this to be reasonable.

Axiom 3 (subadditivity) $\varrho(L_1 + L_2) \le \varrho(L_1) + \varrho(L_2)$ for all $L_1, L_2 \in \mathcal{M}$

Interpr.:

- Reflects the idea of diversification
- Using a non-subadditive ϱ encourages institutions to legally break up into subsidiaries to reduce regulatory capital requirements.
- Subadditivity makes decentralization possible: Assume $L = L_1 + L_2$ and that we want to bound $\varrho(L)$ by M. Choose M_j such that $\varrho(L_j) \le M_j$, $j \in \{1, 2\}$, and $M_1 + M_2 \le M$. Then $\varrho(L) \underset{\text{subadd.}}{\le} \varrho(L_1) + \varrho(L_2) \le M_1 + M_2 \le M$.

Criticism: VaR is ruled out under certain scenarios (see later). VaR is monotone, translation invariant, and positive homogeneous, but in general not subadditive.

Axiom 4 (positive homogeneity) $\varrho(\lambda L) = \lambda\varrho(L)$ for all $L \in \mathcal{M}$, $\lambda > 0$

Interpr.: (or motivation): $\lambda = n \in \mathbb{N}$ and subadditivity imply $\varrho(nL) \le n\varrho(L)$, but n times the same loss L means no diversification, so equality should hold.

Criticism: If $\lambda > 0$ is large, liquidity risk plays a role and one should rather have $\varrho(\lambda L) > \lambda\varrho(L)$ (also to penalize concentration or risk), but this contradicts subadditivity. This has led to convex risk measures (see later), i.e. risk measures ϱ satisfying $\varrho(\lambda L_1 + (1-\lambda)L_2) \le \lambda\varrho(L_1)+(1-\lambda)\varrho(L_2)$ for all $L_1, L_2 \in \mathcal{M}$, $0 \le \lambda \le 1$.

Definition 2.11 (Coherent risk measure)
A risk measure ϱ which satisfies Axioms 1–4 is called *coherent*.

Example 2.12 (Generalized scenario risk measures)

Let $L(\boldsymbol{x})$ denote the hypothetical loss under scenario $\boldsymbol{x}$ (risk-factor change). The generalized scenario risk measure

$$\psi_{\mathcal{X},\boldsymbol{w}}(L) = \max\{\mathbb{E}_{\mathbb{P}}(L(\boldsymbol{X}))\,:\,\boldsymbol{X} \sim \mathbb{P} \in \{\mathbb{P}_1, \dots, \mathbb{P}_n\}\}$$

is coherent. Monotonicity, translation invariance, positive homogeneity are clear (by monotonicity and linearity of $\mathbb{E}(\cdot)$); for subadditivity, note that

$$\begin{aligned}\psi_{\mathcal{X},\boldsymbol{w}}(L_1 + L_2) &= \max\{\underbrace{\mathbb{E}_{\mathbb{P}}(L_1(\boldsymbol{X}) + L_2(\boldsymbol{X}))}_{=\mathbb{E}_{\mathbb{P}}(L_1(\boldsymbol{X}))+\mathbb{E}_{\mathbb{P}}(L_2(\boldsymbol{X}))} : \boldsymbol{X} \sim \mathbb{P} \in \{\mathbb{P}_1, \dots, \mathbb{P}_n\}\} \\ &\le \psi_{\mathcal{X},\boldsymbol{w}}(L_1) + \psi_{\mathcal{X},\boldsymbol{w}}(L_2).\end{aligned}$$

One can show that all coherent risk measures can be represented as generalized scenarios via

$$\varrho(L) = \sup\{\mathbb{E}_{\mathbb{P}}(L) : \mathbb{P} \in \mathcal{P}\}$$

for a suitable set $\mathcal{P}$ of probability measures.

Definition 2.13 (Convex risk measure)
A risk measure ϱ which is monotone, translation invariant and convex is called a *convex risk measure*.

- Justification: Again diversification but they don't have to be positive homogeneous.
- Any coherent risk measure is also a convex risk measure. The converse is not true in general, but for positive homogeneous risk measures, convexity and subadditivity are equivalent.

Theorem 2.14 (Coherence of ES)
ES is a coherent risk measure.

Proof. Monotonicity, translation invariance and positive homogeneity follow from VaR. Subadditivity is more involved but can be shown in various ways; see Embrechts and Wang (2015). $\square$

Superadditivity scenarios for VaR

Under the following scenarios, VaR_α is typically superadditive:

1) L_1, L_2 have skewed distributions;
2) Independent, light-tailed L_1, L_2 and small α;
3) L_1, L_2 have special dependence;
4) L_1, L_2 have heavy tailed distributions.

Exercise 2.15 (Skewed loss distributions)

Consider a portfolio of two independently defaultable zero-coupon bonds (maturity $T = 1$y, nominal/face value 100, paid interest of 5%, default probability $p = 0.009$, no recovery). The loss of bond j (from the lender's/investor's perspective) is thus

$$L_j = \begin{cases} -5, & \text{with prob. } 1-p = 0.991, \\ 100, & \text{with prob. } p = 0.009, \end{cases} \quad j \in \{1,2\}.$$

Set $\alpha = 0.99$. Then $\mathrm{VaR}_\alpha(L_j) = -5$, $j \in \{1,2\}$.

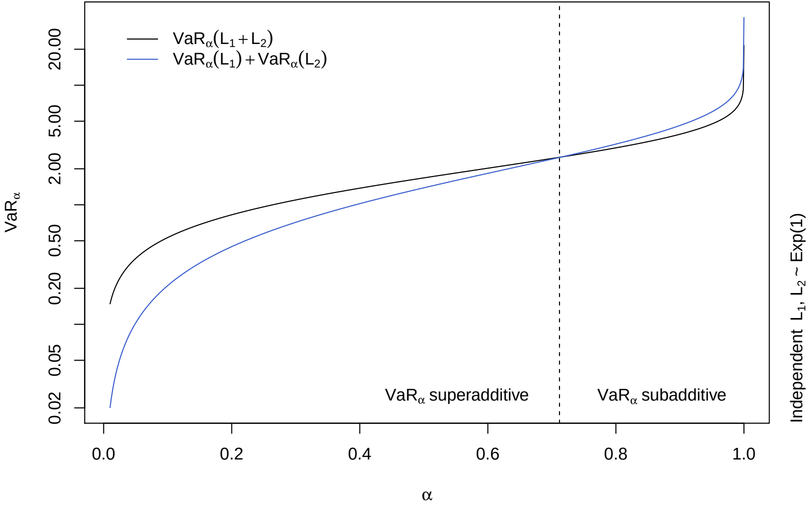
The loss $L_1 + L_2$ is given by

$$L_1 + L_2 = \begin{cases} -10, & \text{with prob. } (1-p)^2 = 0.982081, \\ 95, & \text{with prob. } 2p(1-p) = 0.017838, \\ 200, & \text{with prob. } p^2 = 0.000081. \end{cases}$$

Therefore, $\mathrm{VaR}_\alpha(L_1 + L_2) = 95 > -10 = \mathrm{VaR}_\alpha(L_1) + \mathrm{VaR}_\alpha(L_2)$. Hence VaR_α is superadditive.

Exercise 2.16 (Independent, light-tailed L_1, L_2 and small α)

If $L_1, L_2 \overset{\text{ind.}}{\sim} \mathrm{Exp}(1)$, VaR_α is superadditive $\iff \alpha < 0.71$.



Exercise 2.17 (Special dependence)

Let $\alpha \in (0,1)$, $L_1 \sim \mathrm{U}(0,1)$ and define $L_2 \overset{\text{a.s.}}{=} \begin{cases} L_1, & \text{if } L_1 < \alpha, \\ 1+\alpha-L_1, & \text{if } L_1 \geq \alpha. \end{cases}$

One can show that $L_2 \sim \mathrm{U}(0,1)$. Also, $L_1 + L_2 = \begin{cases} 2L_1, & \text{if } L_1 < \alpha, \\ 1+\alpha, & \text{if } L_1 \geq \alpha, \end{cases}$

from which one can show that

$$F_{L_1+L_2}(x) = \begin{cases} 0, & \text{if } x < 0, \\ x/2, & \text{if } x \in [0, 2\alpha), \\ \alpha, & \text{if } x \in [2\alpha, 1+\alpha), \\ 1, & \text{if } x \geq 1+\alpha. \end{cases}$$

For all $\varepsilon \in \left(0, \frac{1-\alpha}{2}\right)$, we thus obtain that

$$\mathrm{VaR}_{\alpha+\varepsilon}(L_1+L_2) = 1+\alpha \underset{\varepsilon \in \left(0, \frac{1-\alpha}{2}\right)}{>} 2(\alpha+\varepsilon) = \mathrm{VaR}_{\alpha+\varepsilon}(L_1) + \mathrm{VaR}_{\alpha+\varepsilon}(L_2).$$

Exercise 2.18 (Heavy tailed loss distributions)

Let $L_1, L_2 \overset{\text{ind.}}{\sim} F(x) = 1 - x^{-1/2}$, $x \in [1, \infty)$. By deriving the distribution function

$$F_{L_1+L_2}(x) = 1 - 2\sqrt{x-1}/x, \quad x \geq 2,$$

of $L_1 + L_2$ (via the density convolution formula; tedious), one can show (via solving a quadratic equation) that VaR_α is superadditive for all $\alpha \in (0,1)$.

Remark 2.19 (Special case of comonotone risks; elliptical risks)

- In comparison to Exercise 2.17, $L_1 \overset{\text{a.s.}}{=} L_2$ does not lead to the largest $\mathrm{VaR}_\alpha(L_1 + L_2)$ since

$$\mathrm{VaR}_\alpha(L_1 + L_2) \underset{\text{hom.}}{\overset{\text{pos.}}{=}} 2\,\mathrm{VaR}_\alpha(L_1) = \mathrm{VaR}_\alpha(L_1) + \mathrm{VaR}_\alpha(L_2),$$

  so "only" equality.

- VaR_α is subadditive and thus coherent for a certain class of multivariate distributions (strictly including the multivariate normal and t); see later.