Jan Kallsen

Risk Management

Lecture Notes

CAU zu Kiel, WS 16/17, as of October 28, 2022

Contents

Chapter 1

Introduction

These lecture notes are primarily based on [MFE10] and [HL07].

1.1 Motivation and issues

1.1.1 Motivation

Why is risk management important? Banks, financial institutions etc. are subjected to losses. Extremely large losses may lead to bankruptcy and they may also put third parties at risk. Therefore a buffer capital is required from these institutions, units etc., often by law. International standards and guidelines are developped in the Basle committee for banking supervision.

Main issues in this course are:

- How to quantify risk?
- How to measure risk?
- What capital reserve is needed in view of this risk?

1.1.2 Types of risk

Risk here means an event or action which prevents an institution from meeting its obligations or reaching its goals. *Financial risks* typically fall into three groups:

- *Market risk*: risks caused by changing market prices, exchange rates, commodity prices etc.
- *Credit risk*: risk of a lender that the debtor cannot meet his obligations or that the counterparty in a contract cannot meet its obligations.
- *Operational risk*: risk from failure of internal processes, people, systems.

Further risks include e.g. *liquidity risk*, *legal risk*, *reputational risk*. This course concerns primarily market risk.

1.2 Losses and loss operators

1.2.1 Basic concepts

We denote the *portfolio value* at time t as $V(t)$. The *time horizon* is Δt (e.g. a day, ten days, a month, a year). The *profit* in the time interval from t to $t + \Delta t$ is $V(t + \Delta t) - V(t)$. The law of this profit is called *profit and loss (P & L) distribution*. The *loss*

$$L_{[t,t+\Delta t]} := -(V(t + \Delta t) - V(t))$$

is the negative profit. Its law is the *loss distribution*. From now on we consider dates $t_n = n\Delta t$ for $n = 0, 1, 2, \ldots$. As a shorthand notation we write $V_n := V(t_n)$ etc. and

$$L_{n+1} := L_{[n\Delta t,(n+1)\Delta t]} = -(V_{n+1} - V_n).$$

Usually, the present time is denoted by t_n. Our goal is to make a statement on the yet unknown loss L_{n+1} given the data up to time t_n. If the time horizon $k\Delta t$ for risk management is larger than the sampling period Δt of the data, we may also be interested in the law of k-period losses $L_{[t_n,t_n+k\Delta t]} := \sum_{j=1}^{k} L_{n+j}$ given the data up to time t_n.

1.2.2 Modelling

We suppose that profits are a function of risk factors. More specifically,

$$V_n = f(t_n, Z_n)$$

with some known function f and a random vector $Z_n = (Z_{n,1}, \ldots, Z_{n,d})$ of *risk factors* as e.g. asset prices, interest rates, volatility, etc. *Risk factor changes* are denoted by

$$X_{n+1} := Z_{n+1} - Z_n.$$

The loss in period $n + 1$ is a function of X_{n+1} and quantities which are known at t_n. Specifically,

$$\begin{aligned} L_{n+1} &= -(V_{n+1} - V_n) \\ &= -f(t_{n+1}, Z_n + X_{n+1}) + f(t_n, Z_n) \\ &=: \ell_{[n]}(X_{n+1}), \end{aligned}$$

with some function $\ell_{[n]}$. This randomly changing function $\ell_{[n]}$ is called *loss operator*. In general it is nonlinear because the same is true for f.

Taylor's formula provides a useful approximation for small values of X_{n+1}. Indeed,

$$f(t + \tau, z + \zeta) \approx f(t, z) + D_1 f(t, z)\tau + \sum_{i=1}^{d} D_{1+i} f(t, z)\zeta_i,$$

with

$$D_1 f(t,z) := \frac{\partial}{\partial t} f(t,z), \quad D_{1+i} f(t,z) := \frac{\partial}{\partial z_i} f(t,z)$$

yields the *linearised loss*

$$L_{n+1}^{\Delta} := -D_1 f(t_n, Z_n)\Delta t - \sum_{i=1}^{d} D_{1+i} f(t_n, Z_n) X_{n+1,i} \quad (\approx L_{n+1}) \tag{1.1}$$

and, replacing the loss X_{n+1} by x, the *linearised loss operator*

$$\ell_{[n]}^{\Delta}(x) := -D_1 f(t_n, Z_n)\Delta t - \sum_{i=1}^{d} D_{1+i} f(t_n, Z_n) x_i \quad \big(\approx \ell_{[n]}(x)\big)\,.$$

1.2.3 Example of a stock portfolio

Consider a stock portfolio with α_i shares of stock i for $i = 1, \ldots, d$. The price of stock i at time t_n is denoted as $S_{n,i}$, which implies that the portfolio value at time t_n is

$$V_n = \sum_{i=1}^{d} \alpha_i S_{n,i}.$$

As risk factors we consider the logarithmic stock prices

$$Z_{n,i} := \log S_{n,i}.$$

The – in fact dispensable – logarithm turns out to be more convenient for modelling later. The risk factor changes

$$X_{n+1,i} = \log \frac{S_{n+1,i}}{S_{n,i}}, \quad i = 1, \ldots, d$$

are the logarithmic returns of stocks $1, \ldots, d$. The portfolio value can be expressed in terms of risk factors as

$$V_n = \sum_{i=1}^{d} \alpha_i \exp(Z_{n,i}) = f(t_n, Z_n)$$

with $f(t, z_1, \ldots, z_d) := \sum_{i=1}^{d} \alpha_i \exp(z_i)$. The loss and the loss operator are

$$\begin{aligned} L_{n+1} &= -V_{n+1} + V_n \\ &= -\sum_{i=1}^{d} \alpha_i \big(\exp(Z_{n,i} + X_{n+1,i}) - \exp(Z_{n,i}) \big) \\ &= -\sum_{i=1}^{d} \alpha_i S_{n,i} \big(\exp(X_{n+1,i}) - 1 \big) \end{aligned}$$

and

$$\ell_{[n]}(x) = -\sum_{i=1}^{d} \alpha_i S_{n,i} \big(\exp(x_i) - 1 \big),$$

respectively. Since $D_1 f(t_n, Z_{n,1}, \ldots, Z_{n,d}) = 0$ and $D_{1+i} f(t_n, Z_{n,1}, \ldots, Z_{n,d}) = \alpha_i \exp(Z_{n,i}) = \alpha_i S_{n,i}$, the linearised loss resp. loss operator are

$$L_{n+1}^{\Delta} = -\sum_{i=1}^{d} \alpha_i S_{n,i} X_{n+1,i}$$

and

$$\ell_{[n]}^{\Delta}(x) = -\sum_{i=1}^{d} \alpha_i S_{n,i} x_i.$$

1.2.4 Example of a European call option

Consider a portfolio containing a single European call option on a stock. According to the Black-Scholes model, the value of the call and hence the portfolio is of the form

$$C(t, S(t), r, \sigma; T, K), \tag{1.2}$$

where t denotes the present time, $S(t)$ the present stock price, r the riskless interest rate, σ the volatility of the stock, T the maturity of the option, K the strike price, and $C(\ldots)$ the explicitly known function appearing in the Black-Scholes formula. If we consider the stock price, the interest rate, and the volatility as variable, a reasonable vector of risk factors is

$$Z_n = (\log S_n, r_n, \sigma_n),$$

where the index n refers to time t_n as usual. The corresponding risk factor change is

$$X_{n+1} = \left(\log \left(\frac{S_{n+1}}{S_n} \right), r_{n+1} - r_n, \sigma_{n+1} - \sigma_n \right).$$

The loss amounts to

$$\begin{aligned} L_{n+1} &= -C(t_{n+1}, S_{n+1}, r_{n+1}, \sigma_{n+1}; T, K) + C(t_n, S_n, r_n, \sigma_n; T, K) \\ &= -C(t_{n+1}, \exp(Z_{n+1,1}), Z_{n+1,2}, Z_{n+1,3}; T, K) \\ &\quad + C(t_n, \exp(Z_{n,1}), Z_{n,2}, Z_{n,3}; T, K). \end{aligned} \tag{1.3}$$

The linearised loss

$$\begin{aligned} L_{n+1}^{\Delta} &= -D_1 C(t_n, S_n, r_n, \sigma_n; T, K) \Delta t \\ &\quad - D_2 C(t_n, S_n, r_n, \sigma_n; T, K) S_n X_{n+1,1} \\ &\quad - D_3 C(t_n, S_n, r_n, \sigma_n; T, K) X_{n+1,2} \\ &\quad - D_4 C(t_n, S_n, r_n, \sigma_n; T, K) X_{n+1,3} \end{aligned}$$

contains the partial derivatives of the call price function, the so-called Greeks:

$$\begin{aligned} D_1C(t,S,r,\sigma;T,K) &= \frac{\partial}{\partial t}C(t,S,r,\sigma;T,K) \quad \text{“Theta”}, \\ D_2C(t,S,r,\sigma;T,K) &= \frac{\partial}{\partial S}C(t,S,r,\sigma;T,K) \quad \text{“Delta”}, \\ D_3C(t,S,r,\sigma;T,K) &= \frac{\partial}{\partial r}C(t,S,r,\sigma;T,K) \quad \text{“Rho”}, \\ D_4C(t,S,r,\sigma;T,K) &= \frac{\partial}{\partial \sigma}C(t,S,r,\sigma;T,K) \quad \text{“Vega”}. \end{aligned}$$

Equation (1.3) for the loss is less innocent than it may seem. Its validity depends on how volatility σ and possibly even the interest rate r are defined and measured. If σ refers to the standard deviation of stock returns, it is not obvious why option prices should be given by (1.2). Indeed, the Black-Scholes formula relies on assumptions which are violated more or less severely in practice. This problem does not occur with *implied volatility* σ, i.e. the very parameter which makes (1.2) coincide with the call price. In this case, however, one should keep in mind that σ cannot be determined from stock return data and it depends on the option on consideration, i.e. it is typically not the same for different strikes and maturities.

1.3 Quantitative risk management from a bird's eye view

1.3.1 Main steps

In order to apply quantitative risk management, several steps have to be taken.

1. *Exploratory data analysis and modelling.* Since risk management concerns the unknown future, it is typically based on a mathematical model for the loss. The first step is to determine the structure of this model. More specifically, one needs to identify the relevant risk factors $Z_n = (Z_{n,1}, \ldots, Z_{n,d})$, the functional dependence $V_n = f(t_n, Z_n)$, and a stochastic model for the risk factor changes $X_n = Z_n - Z_{n-1}$. These choices are typically based on an exploratory analysis of comparable data from the past as well as on theoretical considerations.

2. *Data collection and parameter estimation.* The model from Step 1 — in particular the one for the risk factor changes — is typically specified only up to some yet unkown parameters. For concrete applications, these must be estimated. To this end, one needs to dispose of reliable data in the first place.

3. *Stochastic forecast.* Based on the now completely specified stochastic model, one can compute an estimate of the conditional law of the future loss L_{n+1} given the data $Z_1, \ldots, Z_n$ up to the present. Possibly, only a quantile, moment etc. is needed instead of the whole law.

4. *Backtesting.* Before these predictions are used in real risk management systems, they should be validated. This is usually done by reviewing whether they would have performed reasonably well in the past.

5. *Draw practical conclusions.* Finally, the prediction or assessment from the model needs to be translated into concrete actions, e.g. concerning buffer capital requirements.

1.3.2 Toy example of a stock portfolio

Let us illustrate these steps in the case of a portfolio consisting only of a share of stock.

1. As in Example 1.2.3 it is natural to consider the stock price S_n itself or, more or less equivalently, its logarithm $Z_n := \log S_n$ as risk factor. A first glance on daily historical data may suggest that *logarithmic stock returns* $X_n := Z_n - Z_{n-1}, n = 1, 2, \ldots$ can be considered as independent and identically distributed (iid) in reasonable approximation. At least, they do not seem to be autocorrelated and they do not vary largely in scale even over long time horizons.

   The simplest choice for the law of the X_n is probably a Gaussian, which implies that the whole model is specified up to two parameters, namely the mean μ and the variance σ^2 of the risk factor changes X_n. The normal distribution can also be justified on theoretical grounds. If we believe that the logarithmic stock price process $\log S(t)$ in continuous time has stationary and independent increments and does not jump, then these increments and hence in particular the Z_n are Gaussian random variables.

2. Given daily past data $X_1, \ldots, X_n$, the natural estimators for μ and σ^2 are

$$\hat{\mu} := \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\sigma}^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2.$$

3. Observe that

$$L_{n+1} = -S_n(e^{X_{n+1}} - 1)$$

   and, more generally,

$$L_{[t_n, t_n + k\Delta t]} = -S_n\left(\exp\left(\sum_{j=1}^{k} X_{n+j}\right) - 1\right)$$

   for the k-period loss. Moreover, S_n is known at time t_n and both X_{n+1} and $\sum_{j=1}^{k} X_{n+j}$ are normally distributed random variables which are independent of the information up to time t_n. Since their parameters are μ, σ^2 resp. $k\mu, k\sigma^2$ in the k-period case, we obtain an estimated loss distribution by inserting the estimates $\hat{\mu}, \hat{\sigma}^2$ for μ, σ.

Sometimes, only quantiles are actually needed, e.g. the 99% quantile of the loss over a ten-day period. The 99% quantile of the standard normal distribution equals $q \approx 2.326$. Consequently,

$$q_L := -S_n \left(e^{10\hat{\mu} - \sqrt{10\hat{\sigma}^2} q} - 1 \right)$$

is a reasonable estimate of the 99% quantile of the loss, i.e. a loss greater than q_L is expected to occur only with probability 1%.

4. Given a long history of stock price data one can examine how often the actual loss surpassed the predicted level q_L. Ideally, this should happen in 1% of the cases. Moreover, these exceedences should be spread more or less evenly over time, see Lemma 1.10 below.

5. If the backtesting produced satisfactory results, one could now impose e.g. q_L as a buffer capital requirement.

1.3.3 Sources of error

In practice, many issues can cause the ultimate assessment of the risk to be faulty.

1. In the modelling step important risk factors may have been overlooked, e.g. counterparty risk, interest rate risk, liquidity risk etc. On top, the functional dependence linking risk factors and portfolio value may not always hold. Recall Example 1.2.4 where the Black-Scholes model is used whose validity in practice is not obvious.

   Finally, the stochastic model for the risk factor changes may not be appropriate to describe real data sufficiently well.

2. Some parametric models require an enormous amount of data for reliable estimation, which may not be available in practice. And even if a long history of data is available, it is not clear whether the model from Step 1 is valid with fixed parameters for such a long time. Structural breaks e.g. after crises may lead to changing parameters and hence error-prone estimates.

3. The forecast may be biased due to an inappropriate linearisation or due to numerical errors in the computation.

4. Backtesting may suffer from the fact that the model is built and tested with the same data. Events that have not occured in the past and are not allowed for in the model either, may still do so in the future.

5. In most cases the buffer capital will not be enough to cover extreme losses. Therefore it should be taken into account how severe consequences turn out to be if things go wrong.

1.3.4 The toy example revisited

1. A closer look on data reveals that logarithmic stock returns seem to be neither independent nor Gaussian. In particular, predictions based on the normal tend to underestimate the probability of big losses dramatically.

2. Since stock return variance changes over time in practice, the estimator $\hat{\sigma}^2$ based on the whole available data may turn out poorly if a forecast for the loss distribution for the next few days is needed. in particular after a sudden increase of volatility it may lead to an underestimation of risks.

3. For risk management the tails of the loss distributions are particularly important. Data suggests that it is precisely this probability of extreme events where the normal distribution fails most drastically. Therefore, the estimate q_L is not going to be very reliable even if it makes sense within the chosen model.

4. Default of the stock and other extreme scenarios can happen in real life, but they are not allowed for in the above toy model. Backtesting does not detect this weakness of the model if these such events did not happen in the available data set.

5. Buffer capital requirements based purely on a fixed quantile (Value-at-Risk) as e.g. q_L ignore the consequences of losses beyond the threshold. Since the probability of such exceedences (e.g. 1%) is not negligible, this is questionable.

1.4 Measuring risks

Risk measurement is performed primarily in order to determine the necessary capital buffer and more generally as a management tool.

1.4.1 Aim

In the end it is desirable to produce a single *number* which

- quantifies the risk and
- yields the necessary capital buffer of a portfolio.

In the following we discuss some examples of such risk measures.

1.4.2 Notional amount

Consider a portfolio whose value equals $V_n = \sum_{i=1}^{d} V_{n,i}$, where $V_{n,i}$ is the amount invested in asset class i, corresponding to loans to central banks, resp. large banks, companies etc. A risk weight α_i of e.g. $0, 0.1, 0.2, 0.5$ or 1 is assigned to any of these asset classes. According

to the standard approch of the Basle committee, the risk figure or regulatory capital assigned to this portfolio equals

$$\varrho := 0.08 \sum_{i=1}^{d} \alpha_i V_{n,i}.$$

This is the amount of equity or buffer capital make the investment in portfolio V_n acceptable. Put differently, a capital reserve of 8% of the risk-weighted investment is required.

The advantages of this risk measure are its simplicity, robustness, and transparency. On the other hand, it does not reflect that diversified portfolios are typically less risky. Compensating risks as e.g. hedges are not recognised either. Altogether, the standard approach represents a rather rough, static classification.

1.4.3 Risk factor sensitivities

If the portfolio value is of the form $V_n = f(t_n, Z_n)$ as in 1.2.2, the risk factor sensitivities

$$\varrho_i := D_{1+i} f(t_n, Z_{n,1}, \ldots, Z_{n,d}), \quad i = 1, \ldots, d$$

represent the dependence of the portfolio value on risk factor changes.

In contrast to the notional amount approach, these risk figures take the dependence on real risks into consideration. On the other hand, they do not provide a single number as desired in 1.4.1. It is not clear at this stage how to obtain a single risk figure from these risk factor sensitivities. Moreover, sensitivities may reflect the risk improperly if f is highly nonlinear or risk factor changes are large.

1.4.4 Scenario-based risk measures

Consider possible values $x_1, \ldots, x_N \in \mathbb{R}^d$ (so-called *scenarios*) of the risk factor change X_{n+1}. Moreover, we attach weights $\omega_1, \ldots, \omega_N \in [0, 1]$ to these scenarios, according to their relevance. If $\ell_{[n]}$ denotes the loss operator of 1.2.2, the corresponding scenario-based risk figure or regulatory capital is defined as

$$\varrho := \max_{i=1,\ldots,N} \omega_i \ell_{[n]}(x_i).$$

The corresponding quantity for k-period losses is defined accordingly.

This relative simple risk measure is obviously robust against estimation errors because it does not involve any estimation at all. It can allow for extreme and even disastrous events that have not yet been observed in the past. On the other hand, the choice of reasonable scenarios is a major challenge for modelling: where do they come from? Are they realistic? Are important ones missing? In any case, "real" risks enter the risk measure only implicitly.

1.4.5 Risk measures based on the loss distribution

This course focuses mainly on risk measures which rely on the law of the loss L_{n+1} (resp. $\sum_{j=1}^{k} L_{n+j}$ for k-period losses). More precisely, denote by P^L the conditional law of L_{n+1} given the information up to time n, i.e. given the observed risk factors $Z_0, \dots, Z_n$ or, put differently, given the σ-field $\mathscr{F}_n := \sigma(Z_0, \dots, Z_n)$. The risk figures resp. regulatory capital requirements ϱ in the following sections (standard deviation, Value-at-Risk, expected shortfall) are a function of P^L or equivalently its cumulative distribution function F_L.

In contrast to the concepts of 1.4.2–1.4.4, these risk measures are based on a stochastic model involving observed data. On the other hand, this does not immediately imply that the forecast performs well. The estimation and even the assumed model may be unreliable. Moreover, it may not be clear whether the future will behave similarly as the past or whether rare extreme events are taken into account appropriately by the model.

1.4.6 Standard deviation

A traditional approach to measure risk is based on the standard deviation of the loss. More specifically,

$$\varrho = c\sqrt{\mathrm{Var}_n(L_{n+1})} = c\sqrt{\int x^2 P^L(dx) - \Big(\int x P^L(dx)\Big)^2}, \tag{1.4}$$

possibly adjusted for the mean by adding $E_n(L_{n+1})$ if the latter is not negligible. Here, $c > 0$ denotes some constant factor,

$$\mathrm{Var}_n(X) := E_n\big((X - E_n(X))^2\big) = E_n(X^2) - (E_n(X))^2$$

the conditional variance, and $E_n(X) := E(X|\mathscr{F}_n)$ the conditional expectation relative to the σ-field $\mathscr{F}_n := \sigma(Z_0, \dots, Z_n)$. As usual, L_{n+1} must be replaced by $\sum_{j=1}^{k} L_{n+j}$ for k-period losses.

The standard deviation is relatively easy to estimate. A less desirable property is its symmetric treatment of profits and losses, which, however, is acceptable if the law P^L is itself symmetric. Moreover, standard deviation obviously does not make sense if the loss does not have finite variance. But more importantly, this risk measure rather quantifies ordinary deviations from the mean as opposed to rare big losses which matter most for risk measurement. Indeed, standard deviation does not tell us much about $P_n(L_{n+1} > x)$ for large x except for the typically rather rough estimate

$$P_n(L_{n+1} > x) \le \frac{\mathrm{Var}_n(L_{n+1})}{(x - E_n(L_{n+1}))^2} = \frac{\varrho^2}{c^2(x - E_n(L_{n+1}))^2},$$

which is based on Chebyshev's inequality. Here $P_n(A) := E_n(1_A)$ denotes the conditional probability of an event A given σ-field $\mathscr{F}_n$. With this notation, we have $P^L = P_n^{L_{n+1}}$, i.e. P_L is the law of L_{n+1} relative to P_n.

1.4.7 Value-at-Risk

The Value-at-Risk is the most widely used risk measure in practice. It is based on quantiles of the loss distribution. We use the same notation as in 1.4.5 and 1.4.6. In particular, we focus on one-period losses L_{n+1}, which could be replaced by k-period losses $\sum_{j=1}^{k} L_{n+j}$ as usual.

Definition 1.1 Let $\alpha \in (0,1)$ be a fixed level. The *Value-at-Risk (VaR)* for level α is defined as

$$\begin{aligned} \mathrm{VaR}_\alpha(L_{n+1}) &= \inf\{\ell \in \mathbb{R} : P_n(L_{n+1} > \ell) \leq 1-\alpha\} \\ &= \inf\{\ell \in \mathbb{R} : 1 - F_L(\ell) \leq 1-\alpha\} \\ &= \inf\{\ell \in \mathbb{R} : F_L(\ell) \geq \alpha\}, \end{aligned}$$

i.e. the smallest number ℓ such that a larger loss occurs only with small probability $1-\alpha$.

Remarks 1.2

1. If the cumulative distribution function F_L is continuous, then
$$P_n(L_{n+1} > \mathrm{VaR}_\alpha(L_{n+1})) = 1-\alpha.$$
2. VaR can be used as a risk measure by setting $\varrho = \mathrm{VaR}_\alpha(L_{n+1})$ for fixed $\alpha \in (0,1)$.
3. In practice, VaR depends both on the level α and on the time horizon. Common time horizons are one resp. 10 days for market risk and one year for credit risk, typical levels are 95%, 99%, 99.9%. E.g. the Basle committee considers the 10-day VaR at level 99% for market risk.
4. VaR is relatively easy to understand and to estimate. It focusus on large losses which makes sense for risk management. However, it lacks *coherence* in the sense of 1.4.9 below. More specifically, diversification may lead to higher risk. Another drawback of VaR is that it takes only the probability but not the size of big losses into account.

The VaR is closely linked to the generalised inverse of F_L.

Definition 1.3

1. For an increasing function $F : \mathbb{R} \to \mathbb{R}$ we call
$$F^{\leftarrow} : \mathbb{R} \to \overline{\mathbb{R}}, \quad F^{\leftarrow}(y) := \inf\{x \in \mathbb{R} : F(x) \geq y\}$$
the *generalised inverse* or *quantile function* of F.
2. If $F : \mathbb{R} \to [0,1]$ is a cumulative distribution function (cdf),
$$q_\alpha(F) := F^{\leftarrow}(\alpha)$$
is called α-*quantile* of F for $\alpha \in (0,1)$.

Lemma 1.4 *1. The quantile function $F^{\leftarrow}$ is increasing and left continuous.*

2. *If F is continuous and strictly increasing, then $F^{\leftarrow} = F^{-1}$, i.e. the generalised inverse is just the ordinary inverse of the invertible function F.*

3. *VaR can be expressed in terms of quantiles as*

$$\mathrm{VaR}_\alpha(L_{n+1}) = q_\alpha(F_L).$$

4. *VaR is* translation invariant *in the sense that*

$$\mathrm{VaR}_\alpha(L_{n+1} + b) = \mathrm{VaR}_\alpha(L_{n+1}) + b, \quad b \in \mathbb{R}.$$

5. *VaR is* positively homogeneous *in the sense that*

$$\mathrm{VaR}_\alpha(aL_{n+1}) = a\mathrm{VaR}_\alpha(L_{n+1}), \quad a \geq 0.$$

Example 1.5 In general, VaR is not *convex* in the sense that

$$\mathrm{VaR}_\alpha\left(\frac{1}{2}(L_{n+1} + \widetilde{L}_{n+1})\right) \leq \frac{1}{2}\big(\mathrm{VaR}_\alpha(L_{n+1}) + \mathrm{VaR}_\alpha(\widetilde{L}_{n+1})\big)$$

does not hold for arbitrary losses $L_{n+1}, \widetilde{L}_{n+1}$. Put differently, diversifying the portfolio appears to increase rather than decrease the risk if the latter is measured in terms of VaR. As an example consider two stochastically independent bonds which may default with probability 0.9%. More specifically, $P_n(L_{n+1} = 1) = 1 - P_n(L_{n+1} = 0) = 0.009$ and likewise for $\widetilde{L}_{n+1}$. The 99% VaR of each amounts to $\mathrm{VaR}_{0.99}(L_{n+1}) = \mathrm{VaR}_{0.99}(\widetilde{L}_{n+1}) = 0$. However, we have $\mathrm{VaR}_{0.99}(\frac{1}{2}(L_{n+1} + \widetilde{L}_{n+1})) = 0.5$, which is larger than the VaR of both constituents.

Example 1.6 If L_{n+1} has conditional law $N(\mu, \sigma^2)$, we can write it as $L_{n+1} = \mu + \sigma X$ for some standard Gaussian X. If

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

denotes the corresponding cdf of X, then $\mathrm{VaR}_\alpha(X) = \Phi^{-1}(\alpha)$ by Lemma 1.4(2, 3). By Statements 3 and 4 of the same lemma, we conclude that

$$\mathrm{VaR}_\alpha(L_{n+1}) = \mu + \sigma\Phi^{-1}(\alpha).$$

Observe that VaR is of standard deviation type (1.4) for centered Gaussian losses.

Example 1.7 If L_{n+1} is of the form $L_{n+1} = s(1 - e^X)$, where X has conditional law $N(\mu, \sigma^2)$, then

$$\begin{aligned} 1 - \alpha &= P_n\left(\frac{X - \mu}{\sigma} < \Phi^{-1}(1 - \alpha)\right) \\ &= P_n\left(X < \sigma\Phi^{-1}(1 - \alpha) + \mu\right) \\ &= P_n\left(s(1 - e^X) > s(1 - \exp(\sigma\Phi^{-1}(1 - \alpha) + \mu))\right) \\ &= P_n\left(L_{n+1} > s(1 - \exp(-\sigma\Phi^{-1}(\alpha) + \mu))\right) \end{aligned}$$

and hence

$$\mathrm{VaR}_\alpha(L_{n+1}) = s(1 - \exp(-\sigma\Phi^{-1}(\alpha) + \mu)). \tag{1.5}$$

Remark 1.8 If X has cdf F instead of having Gaussian conditional law, we obtain

$$\mathrm{VaR}_\alpha(L_{n+1}) = s(1 - \exp(F^{\leftarrow}(1-\alpha)))$$

instead of (1.5).

Example 1.9 Consider a portfolio with $\varphi = 5$ shares of stock. As in 1.2.3 we denote the stock price on day t_n by S_n and assume that the risk factor changes $X_n := \log(\frac{S_n}{S_{n-1}})$, $n = 1, 2, \ldots$ are iid and normally distributed, e.g. with mean $\mu = 0$ and variance $\sigma^2 = 0.02^2 = 0.0004$. According to the previous example, we have

$$\begin{aligned} \mathrm{VaR}_{0.99}(L_{n+1}) &= 5S_n(1 - \exp(-0.02\Phi^{-1}(0.99))) \\ &\approx 5S_n(1 - \exp(-0.046)) \\ &\approx 22.5 \end{aligned}$$

for $S_n = 100$.

If we consider instead the linearised loss $L^{\Delta}_{n+1} = -\varphi S_n X_{n+1}$ with conditional law $N(0, (\varphi S_n \sigma)^2)$, the corresponding VaR equals

$$\begin{aligned} \mathrm{VaR}_{0.99}(L^{\Delta}_{n+1}) &= \varphi S_n \sigma \Phi^{-1}(0.99) \\ &\approx 23 \end{aligned}$$

for $S_n = 100$, which is a good approximation of the true value 22.5 from above.

If, however, a long time horizon is considered, the quality of the approximation decreases. For e.g. 250 trading days, we must multiply σ^2 above by 250, which yields

$$\begin{aligned} \mathrm{VaR}_{0.99}\left(L_{[t_n, t_n+250\Delta t]}\right) &= 5S_n\left(1 - \exp\left(-0.02\sqrt{250}\Phi^{-1}(0.99)\right)\right) \\ &\approx 258 \end{aligned}$$

and the corresponding linearised loss (defined analogously as in (1.1) by first order approximation of the function f in the loss $L_{[t_n, t_n+250\Delta t]} = -f(t_n + 250\Delta t, Z_n + X_{n+1} + \cdots + X_{n+250\Delta t}) + f(t_n, Z_n)$)

$$\begin{aligned} \mathrm{VaR}_{0.99}\left(L^{\Delta}_{[t_n, t_n+250\Delta t]}\right) &= \varphi S_n \sigma \sqrt{250}\Phi^{-1}(0.99) \\ &\approx 364, \end{aligned}$$

which does not appear to be a satisfactory approximation of the true value 258.

The follwing result is relevant for backtesting, see Section 1.3.2. It states that the exceedences of the VaR look like a sequence of coin tosses with success probability $1 - \alpha$.

Lemma 1.10 *The events* $A_{n+1} := \{L_{n+1} > \mathrm{VaR}_\alpha(L_{n+1})\}$, $n = 1, 2, \ldots$ *are independent with* $P(A_{n+1}) = 1 - \alpha$.

Proof. $A_1, \ldots, A_n$ are $\mathscr{F}_n$-measurable by construction. For any $\mathscr{F}_n$-measurable A we have that

$$\begin{aligned} P(A_{n+1} \cap A) &= E(1_{A_{n+1}} 1_A) \\ &= E\big(E(1_{A_{n+1}} 1_A | \mathscr{F}_n)\big) \\ &= E\big(E_n(1_{\{L_{n+1} > \mathrm{VaR}_\alpha(L_{n+1})\}}) 1_A\big) \\ &= E\big((1-\alpha) 1_A\big) \\ &= (1-\alpha) P(A), \end{aligned}$$

which proves the assertion. □

1.4.8 Expected shortfall

We turn now to an alternative to VaR which takes the actual size of large losses into account.

Definition 1.11 Suppose that the conditional law of L_{n+1} given $Z_0, \ldots, Z_n$ is continuous. For fixed level $\alpha \in (0,1)$ the *expected shortfall* at level α is defined as

$$\begin{aligned} \mathrm{ES}_\alpha(L_{n+1}) &:= E_n\big(L_{n+1} \big| L_{n+1} \geq \mathrm{VaR}_\alpha(L_{n+1})\big) \\ &:= \frac{E_n\big(L_{n+1} 1_{\{L_{n+1} \geq \mathrm{VaR}_\alpha(L_{n+1})\}}\big)}{E_n\big(1_{\{L_{n+1} \geq \mathrm{VaR}_\alpha(L_{n+1})\}}\big)}. \end{aligned} \tag{1.6}$$

In other words, the expected shortfall at level α stands for the average loss given that the loss exceeds the VaR at the same level.

The next result indicates how to compute the expected shortfall if the VaR at various levels or the probability density function of the loss is known.

Lemma 1.12 *If $f_L = F_L'$ denotes the probability density function of P^L, we have*

$$\begin{aligned} \mathrm{ES}_\alpha(L_{n+1}) &= \frac{1}{1-\alpha} E_n\big(L_{n+1} 1_{\{L_{n+1} \geq \mathrm{VaR}_\alpha(L_{n+1})\}}\big) \qquad (1.7) \\ &= \frac{1}{1-\alpha} \int_{q_\alpha(F_L)}^{\infty} x f_L(x) dx \\ &= \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_p(L_{n+1}) dp. \qquad (1.8) \end{aligned}$$

Proof. The first two equalities are obvious. For the third we use the substitution $y = F_L(x)$ with $\frac{dy}{dx} = f_L(x)$ yielding

$$\begin{aligned} \int_{q_\alpha}^{\infty} x f_L(x) dx &= \int_{q_\alpha}^{\infty} q_{F_L(x)} f_L(x) dx \\ &= \int_{F_L(q_\alpha)}^{F_L(\infty)} q_y dy \\ &= \int_\alpha^1 \mathrm{VaR}_p(L_{n+1}) dp, \end{aligned}$$

where we use the shorthand q_y for $q_y(F_L)$. □

Remarks 1.13 1. For discrete laws or, more generally, in the presence of atoms the equality $P_n(L_{n+1} \geq \mathrm{VaR}_\alpha(L_{n+1})) = 1-\alpha$ may fail to be true. In this case, we take (1.8) as a definition:

$$\mathrm{ES}_\alpha(L_{n+1}) := \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_p(L_{n+1})dp. \tag{1.9}$$

Instead of (1.7) this *expected shortfall* in the general case satisfies

$$\begin{aligned}\mathrm{ES}_\alpha(L_{n+1}) &= \frac{1}{1-\alpha}E_n\big(L_{n+1}1_{\{L_{n+1}\geq \mathrm{VaR}_\alpha(L_{n+1})\}}\big)\\ &\quad + \mathrm{VaR}_\alpha(L_{n+1})\big(1-\alpha-P_n(L_{n+1}>\mathrm{VaR}_\alpha(L_{n+1}))\big).\end{aligned}$$

2. Since the expected shortfall depends only on the conditional cdf F_L of L_{n+1}, we also write $\mathrm{ES}_\alpha(F_L)$ for $\mathrm{ES}_\alpha(L_{n+1})$.

3. ES_α is translation invariant in the sense that

$$\mathrm{ES}_\alpha(L_{n+1}+b) = \mathrm{ES}_\alpha(L_{n+1}) + b, \quad b\in\mathbb{R}.$$

4. ES_α is positively homogeneous in the sense that

$$\mathrm{ES}_\alpha(aL_{n+1}) = a\mathrm{ES}_\alpha(L_{n+1}), \quad a\geq 0.$$

Example 1.14 If L_{n+1} has conditional law $N(0,1)$, we have

$$\begin{aligned}\mathrm{ES}_\alpha(L_{n+1}) &= \frac{1}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^\infty x\varphi(x)dx\\ &= \frac{1}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^\infty x\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx\\ &= \frac{1}{1-\alpha}\left[-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right]_{\Phi^{-1}(\alpha)}^\infty\\ &= \frac{\varphi(\Phi^{-1}(\alpha))}{1-\alpha}.\end{aligned}$$

If L_{n+1} has conditional law $N(\mu,\sigma^2)$, we can write it as $L_{n+1} = \mu+\sigma X$ for some standard Gaussian X. By translation invariance and positive homogeneity, we conclude that

$$\begin{aligned}\mathrm{ES}_\alpha(L_{n+1}) &= \mathrm{ES}_\alpha(\mu+\sigma X)\\ &= \mu+\sigma\mathrm{ES}_\alpha(X)\\ &= \mu+\sigma\frac{\varphi(\Phi^{-1}(\alpha))}{1-\alpha}.\end{aligned}$$

Example 1.15 If L_{n+1} is of the form $L_{n+1} = s(1 - e^X)$, where X has conditional law $N(\mu, \sigma^2)$, then

$$\mathrm{ES}_\alpha(L_{n+1}) = s\left(1 - \frac{\exp(\mu + \frac{\sigma^2}{2})}{1-\alpha}\Phi\big(-\Phi^{-1}(\alpha) - \sigma\big)\right) \tag{1.10}$$

The proof is left as an exercise.

Remark 1.16 If X in the previous example has cdf F and probability density function (pdf) $f = F'$ instead of having Gaussian conditional law, we obtain instead

$$\mathrm{ES}_\alpha(L_{n+1}) = \frac{1}{1-\alpha}\int_{\mathrm{VaR}_\alpha(L_{n+1})}^{\infty} x f_L(x)dx,$$

where

$$f_L(x) := \frac{f(\log(1 - \frac{x}{s}))}{s - x}$$

is the pdf of L_{n+1}, cf. Lemma 1.12.

1.4.9 Convex and coherent risk measures

Rather than considering concrete risk measures, one may also start with desirable properties and investigate their implications.

Definition 1.17 A mapping $L \mapsto \varrho(L)$ assiging a number to any random variable L is called *convex risk measure* if the following axioms hold:

1. *translation invariance*: $\varrho(L + b) = \varrho(L) + b$ for any loss L and any real number b,
2. *convexity*: $\varrho(\lambda L + (1-\lambda)\widetilde{L}) \leq \lambda\varrho(L) + (1-\lambda)\varrho(\widetilde{L})$ for any losses $L, \widetilde{L}$ and any $\lambda \in [0, 1]$,
3. *monotonicity*: $\varrho(L) \leq \varrho(\widetilde{L})$ if $L \leq \widetilde{L}$,

If we also have

4. *positive homogeneity*: $\varrho(aL) = a\varrho(L)$ for any number $a \geq 0$,

the risk measure is called *coherent*.

As before, $\varrho(L)$ should be thought of the buffer capital that needs to be reserved for L. If the loss is reduced by some fixed amount, the buffer capital can be reduced by the same amount. This natural property is stated above as translation invariance. Convexity reflects the common opinion that diversified portfolios should be considered as less risky. Monotonicity refers to the obvious fact that larger losses should require larger buffer capital. Positive homogeneity means that a risk measure is scale invariant.

How do the risk measures above relate to these axioms? Standard deviation fails to be monotone, so it is not a convex risk measure. VaR satisfies the axoims of translation

invariance, monotonicity, and positive homogeneity. However, convexity does not hold as we observed in Example 1.5.

The following result shows that convex and coherent risk measures can be obtained by considering generalised scenarios.

Lemma 1.18 *Suppose that $Q \mapsto \alpha(Q) \in (-\infty, \infty]$ is a mapping assigning a penalty to any probability measure on Ω. Then*

$$L \mapsto \varrho(L) = \sup\{E_Q(L) - \alpha(Q) : Q \text{ probability measure on } \Omega\} \tag{1.11}$$

defines a convex risk measure.

If $\alpha(Q) \in \{0, \infty\}$ for all Q, then ϱ is in fact coherent. In this case we have

$$\varrho(L) = \sup_{Q \in \mathscr{A}} E_Q(L), \tag{1.12}$$

where $\mathscr{A}$ denotes the set of probability measures Q with $\alpha(Q) = 0$, the so-called generalised scenarios.

Proof. Translation invariance, monotonicity and, in the second case, positive homogeneity are obvious. It remains to show convexity, which follows from

$$\begin{aligned}
&\varrho\Big(\lambda L + (1-\lambda)\widetilde{L}\Big) \\
&= \sup\Big\{E_Q\big(\lambda L + (1-\lambda)\widetilde{L}\big) - \alpha(Q) : Q \text{ probability measure on } \Omega\Big\} \\
&= \sup\Big\{\lambda\big(E_Q(L) - \alpha(Q)\big) + (1-\lambda)\big(E_Q(\widetilde{L}) - \alpha(Q)\big) : Q \text{ probability measure on } \Omega\Big\} \\
&\le \sup\big\{\lambda\big(E_Q(L) - \alpha(Q)\big) : Q \text{ probability measure on } \Omega\big\} \\
&\quad + \sup\Big\{(1-\lambda)\big(E_Q(\widetilde{L}) - \alpha(Q)\big) : Q \text{ probability measure on } \Omega\Big\} \\
&= \lambda\varrho(L) + (1-\lambda)\varrho(\widetilde{L}).
\end{aligned}$$

□

$E_Q(L)$ in Lemma 1.18 can be interpreted as the expected loss under some more or less distorted probabilities. (1.12) means to take the worst of these expected losses. (1.11) is similar, but here the typically nonnegative term $\alpha(Q)$ penalises probability measures acoording to how irrelavent they seem. In this sense $\alpha(Q) = \infty$ means that they are entirely unreasonable and hence do not enter the supremum at all.

Such representations (1.11) resp. (1.12) of convex and coherent risk measures typically exist, more specifically provided that some regularity conditions hold.

The previous result can be used to prove that expected shortfall is a coherent risk measure, at least if we focus on continuous distributions.

Theorem 1.19 *Expected shortfall $L \mapsto \mathrm{ES}_\alpha(L)$ is a coherent risk measure.*

Proof. We show this statement only for the subset of random variables L with continuous distributions.

Consider the set $\mathscr{A}$ of all probability distributions P^B defined via $P^B(A) := P(A|B)$, where B runs through all events with probability $P(B) = 1 - \alpha$. We show that

$$\begin{aligned} \mathrm{ES}_\alpha(L) &= \sup_{Q \in \mathscr{A}} E_Q(L) \\ \Big(&= \frac{1}{1-\alpha} \sup \{E(L1_B) : B \text{ event with } P(B) = 1-\alpha\} \Big), \end{aligned}$$

i.e. ES_α is the coherent risk measure generated by scenario set $\mathscr{A}$. We start by observing that

$$\begin{aligned} (1-\alpha)\mathrm{ES}_\alpha(L) &= \int_\alpha^1 \mathrm{VaR}_p(L)dp \\ &= \int F^{\leftarrow}(p) 1_{[\alpha,1]}(F(F^{\leftarrow}(p)))dp \\ &= E\big(F^{\leftarrow}(U) 1_{[\alpha,1]}(F(F^{\leftarrow}(U))\big) \\ &= E\big(L 1_{[\alpha,1]}(F(L))\big) \\ &= E\big(L 1_{\{L \geq \mathrm{VaR}_\alpha(L)\}}\big) \\ &\leq \sup \{E(L1_B) : B \text{ event with } P(B) = 1-\alpha\}. \end{aligned}$$

where we used that $F^{\leftarrow}(U)$ has the same law as L for uniformly distributed U on $[0,1]$ and, moreover, that $P(L \geq \mathrm{VaR}_\alpha(L)) = 1 - \alpha$.

Now, consider any event B with probability $P(B) = 1 - \alpha$. Note that

$$(L - \mathrm{VaR}_\alpha(L))\big(1_{\{L \geq \mathrm{VaR}_\alpha(L)\}} - 1_B\big) \geq 0.$$

This yields

$$\begin{aligned} E\big(L1_{\{L \geq \mathrm{VaR}_\alpha(L)\}}\big) - E(L1_B) &= E\big((L - \mathrm{VaR}_\alpha(L))\big(1_{\{L \geq \mathrm{VaR}_\alpha(L)\}} - 1_B\big)\big) \\ &\quad + E\big(\mathrm{VaR}_\alpha(L) 1_{\{L \geq \mathrm{VaR}_\alpha(L)\}}\big) \\ &\quad - E(\mathrm{VaR}_\alpha(L) 1_B) \\ &\geq \mathrm{VaR}_\alpha(L)\big(P(L \geq \mathrm{VaR}_\alpha(L)) - P(B)\big) \\ &= 0. \end{aligned}$$

We obtain

$$\begin{aligned} (1-\alpha)ES_\alpha(L) &= E\big(L 1_{\{L \geq \mathrm{VaR}_\alpha(L)\}}\big) \\ &\geq \sup \{E(L1_B) : B \text{ event with } P(B) = 1-\alpha\} \end{aligned}$$

as desired. □

The more intuitive definition (1.6) does not generally lead to a convex risk measure. Indeed, in the situation of Example 1.5 we obtain

$$\begin{aligned}
&E_n\left(\frac{1}{2}(L_{n+1}+\widetilde{L}_{n+1})\middle|\frac{1}{2}(L_{n+1}+\widetilde{L}_{n+1})\geq \mathrm{VaR}_{0.99}\Big(\frac{1}{2}(L_{n+1}+\widetilde{L}_{n+1})\Big)\right)\\
&= E_n\left(\frac{1}{2}(L_{n+1}+\widetilde{L}_{n+1})\middle|\frac{1}{2}(L_{n+1}+\widetilde{L}_{n+1})\geq 0.5\right)\\
&= \frac{11}{21}\\
&\geq 0.1\\
&= \frac{1}{2}\Big(E_n\big(L_{n+1}\big|L_{n+1}\geq 0\big)+E_n\big(\widetilde{L}_{n+1}\big|\widetilde{L}_{n+1}\geq 0\big)\Big)\\
&= \frac{1}{2}\Big(E_n\big(L_{n+1}\big|L_{n+1}\geq \mathrm{VaR}_{0.99}(L_{n+1})\big)+E_n\big(\widetilde{L}_{n+1}\big|\widetilde{L}_{n+1}\geq \mathrm{VaR}_{0.99}(\widetilde{L}_{n+1})\big)\Big).
\end{aligned}$$

It can be shown that ES_α is under some conditions the smallest coherent risk measure dominating VaR_α.

Chapter 2

Common models and methods

2.1 Some basic models

Before we turn to the question how to obtain estimates for VaR and expected shortfall in concrete situations, we consider a number of relatively frequently used models.

2.1.1 Iid Gaussian risk factor changes

In the simplest case the risk factor changes $X_1, X_2, \ldots$ are iid Gaussian random vectors with mean vector $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$. Due to nonlinearity of the loss function, the losses $L_{n+1} = \ell_{[n]}(X_{n+1})$ may no longer be Gaussian, but the linearised losses are of the form

$$L^{\Delta}_{n+1} = \ell^{\Delta}_{[n]}(X_{n+1}) = -(c_n + w_n^\top X_{n+1}) \tag{2.1}$$

with some $c_n \in \mathbb{R}$, $w_n \in \mathbb{R}^d$, which generally depend on the past risk factors $Z_0, \ldots, Z_n$. In other words, L^{Δ}_{n+1} is conditionally Gaussian given observations $Z_0, \ldots, Z_n$, with conditional mean $-(c_n + w_n^\top \mu)$ and conditional variance $w_n^\top \Sigma w_n$.

The advantage of this model is its simplicity. It allows for simple estimation. On the other hand, the linearisation may be inappropriate, in particular for long time horizons. Moreover, the assumption of Gaussianity is also questionable. It often leads to an underestimation of risks.

2.1.2 Iid risk factor changes from a parametric family

If risk factor changes seem to be iid but fail to be described well by the normal distribution, one can try alternative parametric classes of distributions, e.g. the *Student-t distribution* which has heavier tails than the normal. A random variable X is t-distributed with n degrees of freedom and further parameters μ and σ^2 if it is of the form

$$X = \mu + \sigma Z \sqrt{\frac{n}{Z_1^2 + \cdots + Z_n^2}}$$

with independent standard normal random variables $Z, Z_1, \ldots, Z_n$. Its probability density function f is given by

$$f(x+\mu) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi\sigma^2}\Gamma(\frac{n}{2})}\left(1+\frac{(\sigma x)^2}{n}\right)^{-\frac{n+1}{2}}.$$

Such alternative and more flexible families of laws may describe the data better at the cost of possibly more involved computations.

2.1.3 Iid risk factor changes

We may also refrain from any parametric assumption and just assume that risk factor changes are iid. This avoids the model risk of possibly considering an inappropriate family of laws. On the other hand, we obviously cannot use methods from paramatric statistics any more.

2.1.4 Time-varying volatility

Sometimes, even the iid assumption is not warranted by the statistical behaviour of observed risk factor changes. A possible way out is to assume risk factor changes of the form

$$X_n = \sigma_n Y_n, \quad n = 1, 2, \ldots,$$

where the Y_n are iid random variables as discussed above and $(\sigma_n)_{n\in\mathbb{N}}$ denotes a *stochastic volatility* process which varies comparatively slowly over time.

2.1.5 GARCH volatility

For predictions concerning longer time horizons we need a more concrete model. A widely used representative of the *GARCH (generalised autoregressive conditionally heteroscedastic)* class is GARCH(1,1) model, where

$$\begin{aligned} X_n &= \sigma_n Y_n, \\ \sigma_n^2 &= \alpha_0 + \alpha_1 X_{n-1}^2 + \beta\sigma_{n-1}^2 \end{aligned}$$

with parameters $\alpha_0 > 0, \alpha_1 \geq 0, \beta \geq 0$ and iid standard normal random variables Y_n.

2.2 Variance-covariance method

The variance-covariance method relies on the model in 2.1.1, i.e. on iid Gaussian risk factor changes. We assume that risk factors $Z_0, \ldots, Z_n$ have been observed. We suppose that the linear approximation (2.1) is reasonable and that c_n, w_n are known. We are looking for estimates $\widehat{\mathrm{VaR}}_\alpha(L_{n+1}^\Delta), \widehat{\mathrm{ES}}_\alpha(L_{n+1}^\Delta)$ of $\mathrm{VaR}_\alpha(L_{n+1}^\Delta), \mathrm{ES}_\alpha(L_{n+1}^\Delta)$.

The standard estimators for the unknown parameters μ, Σ are

$$\begin{aligned} \hat{\mu}_i &:= \frac{1}{n}\sum_{k=1}^{n} X_{n-k+1,i}, \quad i = 1, \dots, d \\ \hat{\Sigma}_{ij} &:= \frac{1}{n-1}\sum_{k=1}^{n}(X_{n-k+1,i} - \hat{\mu}_i)(X_{n-k+1,j} - \hat{\mu}_j), \quad i, j = 1, \dots, d. \end{aligned}$$

Relying on (2.1) and Examples 1.6, 1.14 we can now compute estimates for VaR and expected shortfall of the loss, namely

$$\widehat{\mathrm{VaR}}_\alpha(L^{\Delta}_{n+1}) = -(c_n + w_n^\top \hat{\mu}) + \sqrt{w_n^\top \hat{\Sigma} w_n}\,\Phi^{-1}(\alpha), \tag{2.2}$$

$$\widehat{\mathrm{ES}}_\alpha(L^{\Delta}_{n+1}) = -(c_n + w_n^\top \hat{\mu}) + \sqrt{w_n^\top \hat{\Sigma} w_n}\,\frac{\varphi(\Phi^{-1}(\alpha))}{1-\alpha}. \tag{2.3}$$

2.3 Historical simulation

In the situation of 2.1.3 we cannot rely on concrete expressions for densities because we did not make any parametric assumptions. Instead we need tools from nonparametric statistics.

2.3.1 Empirical quantiles and expected shortfall

Definition 2.1 Let $X_1, \dots, X_n$ denote independent, identically distributed random variables.

1. The function $F_n : \mathbb{R} \to [0, 1]$,

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} 1_{[X_k,\infty)}(x)$$

is called *empirical distribution function* of $X_1, \dots, X_n$.

2. For $\alpha \in (0, 1)$,

$$q_\alpha(F_n) := F_n^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F_n(x) \geq \alpha\}$$

is the *empirical α-quantile* of F_n.

3.

$$\mathrm{ES}_\alpha(F_n) := \frac{1}{[n(1-\alpha)]+1}\sum_{k=1}^{[n(1-\alpha)]+1} X_{k:n} \tag{2.4}$$

is the *empirical expected shortfall* at level α, where $X_{1:n} \geq X_{2:n} \geq \cdots \geq X_{n:n}$ denote the ordered random variables $X_1, \dots, X_n$ and $[x] := \max\{n \in \mathbb{N} : n \leq x\}$ is the integer part of x.

Remarks 2.2 1. The empirical distribution function is nothing else than the cdf of the *empirical distribution* of $X_1, \ldots, X_n$, i.e. of the (random) probability measure which puts equal probability $1/n$ on any of the observations X_i, $i = 1, \ldots, n$ (unless $X_i = X_j$ for some $i \neq j$, in which case the mass is a multiple of $1/n$).

2. The empirical α-quantile is the α-quantile of the cdf F_n. If $X_i \neq X_j$ for $i \neq j$, we have
$$q_\alpha(F_n) = X_{[n(1-\alpha)]+1:n}, \tag{2.5}$$
i.e. the empirical α-quantile is the $[n(1-\alpha)]+1$-largest observation.

3. Note that the empirical expected shortfall does *not* coincide precisely with the expected shortfall in the sense of (1.9) of a random variable whose law has cdf F_n. Rather, if $X_i \neq X_j$ for $i \neq j$, we have
$$\mathrm{ES}_\alpha(F_n) = E(L|L \geq q_\alpha(L))$$
for a random variable L with cdf F_n, i.e. the empirical expected shortfall is of the form (1.6). Since F_n is a discrete distribution, this is not the expected shortfall of F_n as is observed in Remark 1.13(1). In that sense, this notation clashes with Remark 1.13(2). For large n, however, the difference between (1.6) and (1.9) is small.

It makes sense to use $F_n, q_\alpha(F_n), \mathrm{ES}_\alpha(F_n)$ as estimates for the cdf F of the X_i, its quantiles $q_\alpha(F)$ and expected shortfalls $\mathrm{ES}_\alpha(F)$. Indeed, the following theorem shows that they are typically *consistent* estimators is the sense that they tend asymptotically to the desired limit.

Theorem 2.3 *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with cdf F. By F_n we denote the empirical distribution function of $X_1, \ldots, X_n$.*

1. *We have $F_n(x) \to F(x)$ almost surely for $n \to \infty$ and any $x \in \mathbb{R}$. This holds even uniformly in the sense that*
$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0 \text{ almost surely}$$
for $n \to \infty$ (Glivenko-Cantelli theorem).

2. *If F is strictly increasing, we have*
$$q_\alpha(F_n) \to q_\alpha(F) \text{ almost surely}$$
for $n \to \infty$ and any $\alpha \in (0,1)$.

3. *If F is strictly increasing and continuous, we have*
$$\mathrm{ES}_\alpha(F_n) \to \mathrm{ES}_\alpha(F) \text{ almost surely}$$
for $n \to \infty$ and any $\alpha \in (0,1)$.

Proof.

1. *Step 1:* By the strong law of large numbers we have

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} 1_{[X_k,\infty)}(x) \to E(1_{[X_1,\infty)}(x)) = P(X_1 \le x) = F(x) \quad \text{a.s.}$$

and

$$F_n(x-) = \frac{1}{n}\sum_{k=1}^{n} 1_{(X_k,\infty)}(x) \to E(1_{(X_1,\infty)}(x)) = P(X_1 < x) = F(x-) \quad \text{a.s.}$$

for any fixed x.

Step 2: For $\varepsilon > 0$ choose $-\infty < x_0 < x_1 < \cdots < x_m < \infty$ with $F(x_i-) - F(x_{i-1}) < \varepsilon$ for $i = 1, \dots, m$, where $F(x-) := \lim_{y\uparrow x} F(y)$. By Step 1 there is a (random) n_0 such that $|F_n(x_{i-1}) - F(x_{i-1})| < \varepsilon$ and $|F_n(x_i-) - F(x_i-)| < \varepsilon$ for $i = 1, \dots, m$ and any $n \ge n_0$. For $n \ge n_0$ and any $x \in [x_{i-1}, x_i)$ we have

$$\begin{aligned} |F_n(x) - F(x)| &\le |F_n(x_i-) - F(x_{i-1})| + |F_n(x_{i-1}) - F(x_i-)| \\ &\le |F_n(x_i-) - F(x_i-)| + |F_n(x_{i-1}) - F(x_{i-1})| + 2|F(x_i-) - F(x_{i-1})| \\ &\le 4\varepsilon. \end{aligned}$$

2. For any $x < q_\alpha(F)$ we have $F(x) < \alpha$. Hence $F_n(x) < \alpha$ and therefore $q_\alpha(F_n) \ge x$ for sufficiently large n by Statement 1. Consequently, $\liminf_{n\to\infty} q_\alpha(F_n) \ge x$ a.s.

   Conversely, we have $F(x) > \alpha$ for any $x > q_\alpha(F)$. Hence $F_n(x) > \alpha$ and therefore $q_\alpha(F_n) \le x$ for sufficiently large n by Statement 1. We conclude $\limsup_{n\to\infty} q_\alpha(F_n) \le x$ a.s.

   Together, the claim follows.

3. This is left as an exercise. □

2.3.2 Confidence interval for quantiles

With only finitely many observations we cannot estimate quantiles and other risk measures precisely. Therefore it is desirable to have a confidence interval which contains the unkown true value with high probability.

Definition 2.4 Let $X_1, \dots, X_n$ denote independent, identically distributed random variables with cdf F. Moreover, let $\varrho(F) \in \mathbb{R}$ denote a function of F, typically a risk measure.

We call a (random) interval $(\underline{\varrho}(X_1, \dots, X_n), \overline{\varrho}(X_1, \dots, X_n))$ *confidence interval* at level $p \in (0,1)$ if

$$P\big(\underline{\varrho}(X_1, \dots, X_n) < \varrho(F) < \overline{\varrho}(X_1, \dots, X_n)\big) = p.$$

It is called *centred* if

$$P\big(\underline{\varrho}(X_1, \dots, X_n) \ge \varrho(F)\big) = \frac{1-p}{2} = P\big(\varrho(F) \ge \overline{\varrho}(X_1, \dots, X_n)\big).$$

We now discuss how to come up with a confidence interval at level $p \in (0,1)$ for the α-quantile $q_\alpha(F)$, where F is supposed to be continuous. A natural idea is to choose $(X_{i:n}, X_{j:n})$ for some $i \geq j$, where $X_{1:n}, \ldots, X_{n:n}$ are the ordered observations as in Definition 2.1. Note that $\sum_{k=1}^n 1_{\{X_k \geq q_\alpha(F)\}}$ has a binomial law with parameters n and $1-\alpha$. Consequently, we have

$$\begin{aligned} P(q_\alpha(F) \leq X_{i:n}) &= P\left(\sum_{k=1}^n 1_{\{X_k \geq q_\alpha(F)\}} \geq i\right) \\ &= 1 - F_B(i-1) \end{aligned}$$

and

$$\begin{aligned} P(X_{j:n} \leq q_\alpha(F)) &= 1 - P(q_\alpha(F) < X_{j:n}) \\ &= 1 - P\left(\sum_{k=1}^n 1_{\{X_k > q_\alpha(F)\}} \geq j\right) \\ &= F_B(j-1), \end{aligned}$$

where F_B here denotes the cdf of a binomial law with parameters $n, 1-\alpha$. Looking for a centred confidence interval, we would like to choose i, j such that both expressions equal $\frac{1-p}{2}$, which is generally impossible. As an approximation, choose i such that

$$1 - F_B(i-1) \leq \frac{1-p}{2} < 1 - F_B(i-2)$$

or equivalently

$$F_B(i-2) < \frac{1+p}{2} \leq F_B(i-1). \tag{2.6}$$

If j is chosen such that

$$F_B(j-1) \leq \frac{1-p}{2} < F_B(j), \tag{2.7}$$

we obtain

$$P(X_{i:n} < q_\alpha(F) < X_{j:n}) = -F_B(j-1) + F_B(i-1) =: p' \geq p,$$

i.e. $(X_{i:n}, X_{j:n})$ is a confidence interval at level $p' \geq p$. For large n, the difference $p' - p$ tends to 0.

Example 2.5 In the above setup consider $n = 10$ observations. We look for a confidence interval at level at least $p = 0.75$ for $q_{0.8}(F)$. If F_B denotes the cdf of the binomial law with parameters 10 and 0.2, (2.6) and (2.7) hold for $i = 4$ and $j = 1$, respectively. In this case, the level p' of the confidence interval is

$$p' = -F_B(0) + F_B(3) \approx 0.772.$$

2.3.3 Application to risk management

In the situation of 1.2.2 we assume that the risk factor changes $X_1, X_2, \ldots$ are iid as in 2.1.3. We are looking for estimates $\widehat{\mathrm{VaR}}_\alpha(L_{n+1}), \widehat{\mathrm{ES}}_\alpha(L_{n+1})$ of $\mathrm{VaR}_\alpha(L_{n+1}), \mathrm{ES}_\alpha(L_{n+1})$, where $L_{n+1} = \ell_{[n]}(X_{n+1})$ as in 1.2.2. The idea is to replace the unknown law of the X_i with the empirical distribution and, consequently, replace the unknown conditional law of $\ell_{[n]}(X_{n+1})$ by the empirical law of $\ell_{[n]}(X_{n-k+1})$, $k = 1, \ldots, n$ in order to obtain empirical quantiles and expected shortfalls as in 2.3.1. More specifically, we set $\ell_k := \ell_{[n]}(X_{n-k+1})$, $k = 1, \ldots, n$ and denote the corresponding ordered observations as $\ell_{1:n} \geq \cdots \geq \ell_{n:n}$. Motivated by (2.5) and (2.4), we define estimates

$$\widehat{\mathrm{VaR}}_\alpha(L_{n+1}) := \hat{q}_\alpha(L_{n+1}) := \ell_{[n(1-\alpha)]+1:n}$$

and

$$\widehat{\mathrm{ES}}_\alpha(L_{n+1}) := \frac{1}{[n(1-\alpha)]+1} \sum_{k=1}^{[n(1-\alpha)]+1} \ell_{k:n}.$$

Remarks 2.6

1. In practice, the number of past observations n should be chosen with care. Even if a long record of data is available, it may be preferable to use only the recent past. Indeed, it is not obvious whether stationarity really holds over long time horizons. Using too few observations, however, leads to unreliable estimates because of their large variance.

2. The above estimates are easy to compute and do not rely on parametric assumptions that may not hold. In particular, any dependence between the components of multivariate observations X_i is reflected in the estimates. As a drawback, large data sets are needed to come up with reliable estimates for the tails and hence for VaR and expected shortfall. In particular, the greatest expected loss never exceeds the largest loss observed in the past.

2.4 Maximum likelihood estimation

In 2.1.2 or 2.1.5, the model is given up to a finite number of parameters. In this situation one uses methods from parametric statistics to estimate the parameters. We consider here maximum likelihood estimation because it often has desirable properties. The *maximum likelihood estimator (MLE)* of a parameter vector is the value such that the density is maximal.

Suppose that the random vector $(X_1, \ldots, X_n)$ has multivariate pdf $(x_1, \ldots, x_n) \mapsto f_\vartheta^{(n)}(x_1, \ldots, x_n)$ (relative to Lebesgue measure in $\mathbb{R}^n$). Here, $\vartheta \in \mathbb{R}^d$ denotes the unkown parameter vector. The maximum likelihood estimate $\hat{\vartheta}$ is the parameter ϑ maximizing the density or, equivalently, the *log likelihood* $\vartheta \mapsto \log(f_\vartheta^{(n)}(X_1, \ldots, X_n))$, which obviously depends on the observations $X_1, \ldots, X_n$.

From asymptotic statistics it is known that the MLE $\hat{\vartheta}$ often tends to the true parameter vector ϑ when the number of observations tends to infinity *(consistency)*. Moreover, it is often *asymptotically normal*, i.e. the law of the standardised MLE converges weakly to a normal distribution for $n \to \infty$.

If the random variables $X_1, \ldots, X_n$ are iid with pdf f_ϑ, then we have

$$\log(f_\vartheta^{(n)}(x_1, \ldots, x_n)) = \sum_{i=1}^{n} \log(f_\vartheta(x_i)).$$

More generally, if the conditional law of X_i given $X_1, \ldots, X_{i-1}$ has pdf

$$x \mapsto f_{\vartheta,i}(X_1, \ldots, X_{i-1}; x),$$

the joint log likelihood equals

$$\log(f_\vartheta^{(n)}(x_1, \ldots, x_n)) = \sum_{i=1}^{n} \log(f_{\vartheta,i}(x_1, \ldots, x_{i-1}; x_i)).$$

The estimated law can now be used to come up with estimates of VaR and expected shortfall, either by computing quantiles and partial moments directly or by Monte Carlo simulation, which is discussed below.

Example 2.7 In the GARCH(1,1) model of 2.1.5 we consider $\vartheta := (\alpha_0, \alpha_1, \beta, \sigma_0)$ as parameter vector. Observe that the conditional law of X_i given $X_1, \ldots, X_{i-1}$ is Gaussian with mean 0 and variance σ_n, i.e.

$$\log(f_{\vartheta,i}(X_1, \ldots, X_{i-1}; x)) = -\frac{x^2}{2\sigma_i^2} - \frac{1}{2}\log(2\pi\sigma_i^2).$$

Here σ is a function of ϑ and $X_1, \ldots, X_{i-1}$, defined recursively by

$$\sigma_n^2 := \alpha_0 + \alpha_1 X_{n-1}^2 + \beta\sigma_{n-1}^2.$$

for $n = 1, 2, \ldots$.

2.5 Monte Carlo methods

Recall that the loss under consideration is of the form $L_{n+1} = \ell_{[n]}(X_{n+1})$. As usual, we are interested to compute $\mathrm{VaR}_\alpha(L_{n+1}) = \mathrm{VaR}_\alpha(\ell_{[n]}(X_{n+1}))$ resp. $\mathrm{ES}_\alpha(L_{n+1}) = \mathrm{ES}_\alpha(\ell_{[n]}(X_{n+1}))$. In principle this is possible if the conditional law of X_{n+1} given past data is known. In concrete models, however, a closed-form expression for these quantiles or partial moments is often not available. But in many situations one knows how to simulate random variables whose distribution equals the conditional law of X_{n+1}. The idea is to use the estimates from 2.3.3, but with simulated data instead of the historical observations.

This approach is quite flexible. It only requires that we can simulate from the model, which may even be the case if risk factor changes are not independent or stationary. By increasing N we can get in principle arbitrary precision. However, achieving high precision is computationally costly. Moreover, unlike the situation of 2.1.3 resp. 2.3.3 parameters still need to be estimated beforehand, e.g. by maximum likelihood as in Section 2.4.

2.5.1 How to proceed

Simulate N independent random numbers $x_1, \ldots, x_N$ drawn from the conditional law of X_{n+1} given the past observations $Z_0, \ldots, Z_n$. Now use the empirical law of these simulations $x_1, \ldots, x_N$ instead of real observations in order to come up with estimates $\widehat{\mathrm{VaR}}_\alpha(L_{n+1}), \widehat{\mathrm{ES}}_\alpha(L_{n+1})$ as in 2.3.3. More specifically, set

$$\widehat{\mathrm{VaR}}_\alpha(L_{n+1}) := \hat{q}_\alpha(L_{n+1}) := \ell_{[N(1-\alpha)]+1:N}$$

and

$$\widehat{\mathrm{ES}}_\alpha(L_{n+1}) := \frac{1}{[N(1-\alpha)]+1} \sum_{k=1}^{[N(1-\alpha)]+1} \ell_{k:N},$$

where $\ell_k := \ell_{[n]}(x_k)$, $k = 1, \ldots, N$ and $\ell_{1:N} \geq \cdots \geq \ell_{N:N}$ are the same numbers ordered by size.

If the value at risk or expected shortfall for m instead of one period ahead is needed, one proceeds along the same lines, but of course one needs to simulate from the conditional law of X_{n+m} given $Z_0, \ldots, Z_n$

Chapter 3

Extreme value theory in risk measurement

Extreme value theory (EVT) provides mathematical methods for statements on probabilities and the statistics of extreme events. They are of interest here because the VaR and expected shortfall for large α concern precisely events of such type.

3.1 Extreme value theory

3.1.1 Motivation

1. If one uses the empirical distribution for estimating the probability of large losses, one faces a major problem. Since extremely large losses have been observed rarely — if at all — the empirical law does not yield a reasonable estimate.

2. We focus on laws with *heavy tails*, which here means that the *survival function* $\overline{F}(x) = 1 - F(x)$ of the cdf F satisfies

$$\lim_{x\to\infty} \frac{\overline{F}(x)}{e^{-\lambda x}} = \infty, \qquad \text{for all } \lambda > 0.$$

   Put differently, the law assigns more mass to extreme events than any exponential distribution. But note that *heavy tailed* does not always refer to the same property in the literature. Sometimes, the law of a random variable X is instead called "heavy tailed" if $E(|X|^n) < \infty$ does not hold for arbitrarily large n.

3. In order to assess whether the data is heavy tailed it makes sense to start with an exploratory data analysis based on quantile-quantile plots.

3.1.2 Quantile-quantile plot (qq plot)

Suppose that $X_1, \ldots, X_n$ are iid random variables with order statistics $X_{1:n} \geq X_{2:n} \cdots \geq X_{n:n}$. Moreover, let F be an arbitrary cdf, which is to be interpreted as a candidate for the

unknown true cdf F_X of the X_i. The *quantile-quantile plot (qq plot)* of the data $X_1, \dots, X_n$ against F is the graphical representation of the set of points

$$\{(X_{k:n}, F^{\leftarrow}(\tfrac{n-k+1}{n+1})) : k = 1, \dots, n\}.$$

The first coordinate of any of these points is an empirical quantile, the second one the corresponding quantile of F. If $F = F_X$, then we have $F(X_{k:n}) \approx F_n(X_{k:n}) = \frac{n-k+1}{n+1}$ and hence $X_{k:n} \approx F^{\leftarrow}(\frac{n-k+1}{n+1})$, where F_n denotes the empirical cdf of $X_1, \dots, X_n$. Put differently, the points should be close to the line $\{(x, x) : x \in \mathbb{R}\}$.

If we have at least $F_X(x) = F(\frac{x-\mu}{\sigma})$ for some $\mu \in \mathbb{R}$, $\sigma > 0$, then X has cdf F up to some shift and rescaling. In this case

$$F^{\leftarrow}(\tfrac{n-k+1}{n+1}) = \tfrac{F_X^{\leftarrow}(\frac{n-k+1}{n+1})-\mu}{\sigma},$$

i.e. the points are still close to a line, namely to $\{(x, (x-\mu)/\sigma) : x \in \mathbb{R}\}$.

If the plot is curved downwards at the left or upwards at the right, this indicates that F has heavier tails than the data. The opposite holds if F has lighter tails than the data.

3.1.3 Regular variation

The heaviness of the tails can be quantified.

Definition 3.1 1. A function $h : (0, \infty) \to (0, \infty)$ is called *regularly varying in ∞ with index $\varrho \in \mathbb{R}$* (written $h \in \mathrm{RV}_\varrho$) if

$$\lim_{t\to\infty} \frac{h(tx)}{h(t)} = x^\varrho \quad \text{for any } x > 0.$$

For $\varrho = 0$ we say that h is *slowly varying in ∞* (and often use letter L rather than h).

2. A random variable X with cdf F is called *regularly varying* if $\overline{F} \in \mathrm{RV}_{-\alpha}$ for some $\alpha \geq 0$, where $\overline{F}(x) = 1 - F(x)$.

Remarks 3.2 1. $h \in \mathrm{RV}_\varrho$ implies that $h(x) = L(x)x^\varrho$ for some $L \in \mathrm{RV}_0$.

2. $h \in \mathrm{RV}_\varrho$ for $\varrho < 0$ implies

$$\lim_{t\to\infty} \sup_{x\in[b,\infty)} \left| \frac{h(tx)}{h(t)} - x^\varrho \right| = 0 \quad \text{for any } b > 0,$$

i.e. the convergence is uniform on intervals $[b, \infty)$.

3. Examples for slowly varying functions are $L(x) = c$ with constant $c > 0$, any function L with $\lim_{x\to\infty} L(x) = c \in (0, \infty)$, or $L(x) = \log(1 + x)$, respectively. It may occur that both $\liminf_{x\to\infty} L(x) = 0$ and $\limsup_{x\to\infty} L(x) = \infty$, i.e. oscillations between 0 and ∞ are possible. Moreover, $x^{-\varrho}L(x) \to 0$, $x^\varrho L(x) \to \infty$ for $x \to \infty$ and $\varrho > 0$.

4. An example for cdfs F with $\overline{F} \in \mathrm{RV}_{-\alpha}$ is $F(x) = 1_{[1,\infty)}(x)(1 - x^{-\alpha})$ because

$$\frac{\overline{F}(tx)}{\overline{F}(t)} = \frac{(tx)^{-\alpha}}{t^{-\alpha}} = x^{-\alpha}.$$

5. If $X \geq 0$ is regularly varying with $\overline{F} \in \mathrm{RV}_{-\alpha}$, then $E(X^\beta) < \infty$ for $\beta < \alpha$ and $E(X^\beta) = \infty$ for $\beta > \alpha$. The converse is generally not true. But nevertheless, "X is regularly varying" means more or less that its moments cease to exist for large β. Moreover, the parameter α indicates the heaviness of the tails of the law of X.

Example 3.3 Let random variables $X, Y \geq 0$ stand for losses of an insurance company. X refers to fire insurance and has cdf F such that $\overline{F} \in \mathrm{RV}_{-\alpha}$. Moreover, Y refers to car insurance and is supposed to satisfy $E(Y^k) < \infty$ for any $k > 0$. We want to determine $\lim_{x\to\infty} P(X > x|X+Y > x)$, i.e. the probability the loss due to fire is large given that the overall loss is large.

To this end, fix $\varepsilon \in (0, 1)$ and $x > 0$. Then

$$\begin{aligned} P(X+Y>x) &= P(X+Y>x, X>(1-\varepsilon)x) + P(X+Y>x, X\leq(1-\varepsilon)x) \\ &\leq P(X+Y>x, X>(1-\varepsilon)x) + P(X+Y>x, Y>\varepsilon x) \\ &\leq P(X>(1-\varepsilon)x) + P(Y>\varepsilon x) \end{aligned}$$

which implies

$$\begin{aligned} 1 &\leq \frac{P(X+Y>x)}{P(X>x)} \\ &\leq \frac{P(X>(1-\varepsilon)x)}{P(X>x)} + \frac{P(Y>\varepsilon x)}{P(X>x)} \\ &\leq \frac{P(X>(1-\varepsilon)x)}{P(X>x)} + \frac{E(Y^{2\alpha})}{(\varepsilon x)^{2\alpha}P(X>x)} \qquad (3.1) \\ &\overset{x\to\infty}{\longrightarrow} (1-\varepsilon)^{-\alpha} + 0. \qquad (3.2) \end{aligned}$$

In (3.1) we need the Markov inequality and in (3.2) that $P(X > x) = x^{-\alpha}L(x)$ for some slowly varying L. Since ε can be chosen arbitrarily small, we obtain

$$\lim_{x\to\infty} \frac{P(X+Y>x)}{P(X>x)} = 1$$

and hence

$$\begin{aligned} \lim_{x\to\infty} P(X>x|X+Y>x) &= \lim_{x\to\infty} \frac{P(X>x, X+Y>x)}{P(X+Y>x)} \\ &= \lim_{x\to\infty} \frac{P(X>x)}{P(X+Y>x)} = 1. \end{aligned}$$

Consequently, if a large total loss occurs, it is probably due to fire.

3.2 Hill estimator

The goal in this section is to estimate the parameter α of a regularly varying random variable X.

3.2.1 Derivation

Suppose that $X_1, \ldots, X_n$ are iid regularly varying random variables, i.e. $\overline{F}(x) = x^{-\alpha}L(x)$, where F is the cdf of the X_i, moreover $\overline{F} = 1 - F$, $\alpha > 0$, L slowly varying. As a tool we use *Karamata's theorem*, which says that

$$\int_u^\infty x^\beta L(x)dx \sim -\frac{u^{\beta+1}L(u)}{\beta+1} \quad \text{for } u \to \infty$$

for $\beta < -1$ and slowly varying L. We write here $f \sim g$ for $f/g \to 1$. (This result is obvious for constant L).

This yields

$$\begin{aligned}
&\frac{1}{\overline{F}(u)} \int_u^\infty (\log(x) - \log(u))dF(x) \\
&= \frac{1}{\overline{F}(u)} \left(\left[-(\log(x) - \log(u))\overline{F}(x) \right]_u^\infty + \int_u^\infty \frac{\overline{F}(x)}{x} dx \right) \\
&= \frac{1}{u^{-\alpha}L(u)} \int_u^\infty x^{-\alpha-1}L(x)dx \\
&\overset{u\to\infty}{\longrightarrow} \frac{1}{\alpha},
\end{aligned}$$

where we use integration by parts in the first equality.

In order to estimate α, we replace the unknown cdf F by its empirical counterpart

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{[X_i,\infty)}(x).$$

Moreover, u is chosen as $X_{k:n}$ for some small k. Since u is large, we obtain

$$\begin{aligned}
\frac{1}{\alpha} &\approx \frac{1}{\overline{F}(u)} \int_u^\infty (\log(x) - \log(u))dF(x) \qquad (3.3) \\
&\approx \frac{1}{\overline{F}_n(X_{k:n})} \int_{X_{k:n}}^\infty (\log(x) - \log(X_{k:n}))dF_n(x) \\
&= \frac{n}{k-1} \sum_{j=1}^{k-1} (\log(X_{j:n}) - \log(X_{k:n})) \frac{1}{n} \\
&\approx \frac{1}{k} \sum_{j=1}^{k-1} (\log(X_{j:n}) - \log(X_{k:n})). \qquad (3.4)
\end{aligned}$$

If $k = k(n)$ satisfies both $k \to \infty$ and $k/n \to 0$, we expect convergence of (3.4) to $1/\alpha$ for $n \to \infty$ because then the number of summand $X_{k:n}$ goes to ∞ as well. This motivates the *Hill estimator*

$$\hat{\alpha}_{k,n}^{(H)} := \frac{k}{\sum_{j=1}^{k-1}(\log(X_{j:n}) - \log(X_{k:n}))},$$

which we expect to converge to α if both $k \to \infty$ and $k/n \to 0$.

3.2.2 Choice of the the number k

In practice it is not evident how to choose the number k of upper order statistics. If it is chosen too large, we obtain a bias because the approximation (3.3) generally holds only for large u and hence small k. But if k is chosen too small, we average over too few data points which leads to high variance.

The idea is to choose k graphically by looking at the *Hill plot*, i.e. the graphical representation of the set of points

$$\left\{(k, \hat{\alpha}_{k,n}^{(H)}) : k = 2, \ldots, n\right\}.$$

If we are lucky, the plot looks approximately for a while constant after some initial oscillations. This is the part where k and hence the estimate should be chosen.

3.2.3 Estimation of VaR

In risk management applications, we want to estimate e.g. $\mathrm{VaR}_p(F) = q_p(F)$ for large p. If $\overline{F}$ is regularly varying with index $-\alpha$ and $X_{k:n}$ is sufficiently large, we approximate

$$\begin{aligned}
\overline{F}(x) &= \overline{F}\left(\frac{x}{X_{k:n}} X_{k:n}\right) \\
&\approx \left(\frac{x}{X_{k:n}}\right)^{-\alpha} \overline{F}(X_{k:n}) \\
&\approx \left(\frac{x}{X_{k:n}}\right)^{-\hat{\alpha}_{k,n}^{(H)}} \overline{F}_n(X_{k:n}) \\
&= \left(\frac{x}{X_{k:n}}\right)^{-\hat{\alpha}_{k,n}^{(H)}} \frac{k-1}{n} \\
&\approx \frac{k}{n}\left(\frac{x}{X_{k:n}}\right)^{-\hat{\alpha}_{k,n}^{(H)}}.
\end{aligned}$$

This suggests to use the estimator

$$\widehat{\overline{F}(x)} := \frac{k}{n}\left(\frac{x}{X_{k:n}}\right)^{-\hat{\alpha}_{k,n}^{(H)}}.$$

Since $q_p(F) = \inf\{x \in \mathbb{R} : \overline{F}(x) \leq 1-p\}$, this in turn leads to the estimator

$$\begin{aligned}
\widehat{q_p(F)} &= \inf\left\{x \in \mathbb{R} : \widehat{\overline{F}(x)} \leq 1-p\right\} \\
&= \inf\left\{x \in \mathbb{R} : \frac{k}{n}\left(\frac{x}{X_{k:n}}\right)^{-\hat{\alpha}_{k,n}^{(H)}} \leq 1-p\right\} \\
&= \left(\frac{n}{k}(1-p)\right)^{-1/\hat{\alpha}_{k,n}^{(H)}} X_{k:n} \qquad (3.5)
\end{aligned}$$

for $q_p(F)$. Using (3.5) we obtain a reasonable estimate for VaR at levels α where no empirical data is available.

3.2.4 Estimation of expected shortfall

Inserting the estimated VaR (3.5) in (1.8), we obtain the following estimate for the expected shortfall:

$$\begin{aligned}
\widehat{\mathrm{ES}}_p(F) &= \frac{1}{1-p}\int_p^1 \widehat{q_r(F)}dr \\
&= \frac{1}{1-p}\left[-\left(1-1/\hat{\alpha}_{k,n}^{(H)}\right)^{-1}(1-r)^{1-1/\hat{\alpha}_{k,n}^{(H)}}\right]_p^1 \left(\frac{n}{k}\right)^{-1/\hat{\alpha}_{k,n}^{(H)}} X_{k:n} \\
&= \left(1-1/\hat{\alpha}_{k,n}^{(H)}\right)^{-1}\left(\frac{n}{k}(1-p)\right)^{-1/\hat{\alpha}_{k,n}^{(H)}} X_{k:n} \\
&= \left(1-1/\hat{\alpha}_{k,n}^{(H)}\right)^{-1}\widehat{q_p(F)}.
\end{aligned}$$

3.3 Peaks over threshold (POT) method

In this section we discuss an alternative approach to estimate $\overline{F}(x)$, $\mathrm{VaR}_p(F)$, $\mathrm{ES}_p(F)$ for large x, p.

3.3.1 Motivation

Let X be a random variable with cdf F. Define its *excess distribution function* over threshold u as

$$F_u(x) := P(X - u \leq x | X > u), \quad x \geq 0.$$

If X is exponentially distributed with cdf F, then $F_u = F$ for any $u \geq 0$.

Now let X instead be a regularly varying random variable with parameter $\alpha > 0$. Then

$$\begin{aligned} \overline{F}_u(x) &:= 1 - F_u(x) \\ &= \frac{P(X > x + u)}{P(X > u)} \\ &= \frac{\overline{F}(u + x)}{\overline{F}(u)} \qquad (3.6) \\ &= \frac{\overline{F}(u(1 + x/u))}{\overline{F}(u)}. \qquad (3.7) \end{aligned}$$

$\overline{F} \in \mathrm{RV}_{-\alpha}$ and Remark 2 in 3.1.3 yield

$$\lim_{u\to\infty} \sup_{\lambda\geq 1} \left| \frac{\overline{F}(\lambda u)}{\overline{F}(u)} - \lambda^{-\alpha} \right| = 0.$$

In view of (3.7) with $\lambda = 1 + x/u$ we obtain

$$\lim_{u\to\infty} \sup_{x\geq 0} \left| \overline{F}_u(x) - \left(1 + \frac{x}{u}\right)^{-\alpha} \right| = 0.$$

The second term on the left is the survival function of some cdf $G_{\frac{1}{\alpha},\frac{u}{\alpha}}$, namely the generalised Pareto distributions with parameters $\frac{1}{\alpha}, \frac{u}{\alpha}$. The *generalised Pareto distribution (GPD)* with parameters $\gamma > 0$, $\beta > 0$ is defined by its cdf

$$G_{\gamma,\beta}(x) = 1 - \left(1 + \frac{\gamma x}{\beta}\right)^{-\frac{1}{\gamma}}, \quad x \geq 0.$$

Consequently, we have

$$\lim_{u\to\infty} \sup_{x>0} \left| \overline{F}_u(x) - \overline{G}_{\gamma,\beta(u)}(x) \right| = 0$$

with $\gamma = 1/\alpha$, $\beta(u) = u/\alpha$.

In other words, the tails of a regularly varying random variable with parameter $\alpha > 0$ resemble a generalised Pareto distribution. This fact is used in 3.3.5 and 3.3.6 to estimate VaR and expected shortfall.

3.3.2 Parameter estimation

We assume that $X_1, \ldots, X_n$ are iid random variables with cdf F such that

$$\overline{F}_u \approx \overline{G}_{\gamma,\beta} \qquad (3.8)$$

holds for some sufficiently large u, where $\gamma > 0, \beta > 0$ are some parameters and F_u denotes the excess distribution function of the X_i. By 3.3.1 this holds e.g. for regularly varying random variables with parameter $\alpha > 0$. Our goal is to estimate γ and β.

Denote by

$$N_u := |\{i \in \{1, \ldots, n\} : X_i > u\}|$$

the number of exceedences of u by the observations $X_1, \ldots, X_n$. Moreover, the excesses $Y_1, \ldots, Y_{N_u}$ are defined as $Y_i := X_j - u$, where X_j denotes the i'th observation exceeding u. These excesses $Y_1, \ldots, Y_{N_u}$ are iid and their cdf is approximately $G_{\gamma,\beta}$ by (3.8).

We estimate the parameters γ, β by maximum likelihood as in Section 2.4. Since the pdf of the Y_i equals

$$g_{\gamma,\beta}(y) = G'_{\gamma,\beta}(y) = \frac{1}{\beta}\left(1 + \frac{\gamma y}{\beta}\right)^{-\frac{1}{\gamma}-1},$$

the joint likelihood of $Y_1, \ldots, Y_{N_j}$ is given by

$$L(\gamma, \beta; y_1, \ldots, y_{N_u}) = \prod_{i=1}^{N_u} g_{\gamma,\beta}(y_i).$$

Given the observations $Y_1, \ldots, Y_{N_u}$ of the Y_i, the MLE $\hat{\gamma}, \hat{\beta}$ can now be determined numerically by maximising

$$\log L(\gamma, \beta; Y_1, \ldots, Y_{N_u}) = -N_u \log(\beta) - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{N_u} \log\left(1 + \frac{\gamma}{\beta} Y_i\right)$$

as a function of γ, β.

As mentioned in Section 2.4, the MLE is approximately Gaussian for large N_u, more specifically, $(\hat{\gamma} - \gamma, \frac{\hat{\beta}}{\beta} - 1)$ is approximately normal with mean vector $(0, 0)$ and covariance matrix Σ^{-1}/N_u, where

$$\Sigma^{-1} = (1 + \gamma) \begin{pmatrix} 1 + \gamma & -1 \\ -1 & 2 \end{pmatrix}.$$

Similarly as the number k in 3.2.2, it is not obvious how to choose the threshold u. For small u, it is not obvious whether the approximation (3.8) makes sense at all. For large u, however, the variance of the MLE may be large because it involves only few observations. A tool to decide how to choose u is the mean excess plot discussed below.

3.3.3 Mean-excess plot

For random variables X with $E(X) < \infty$ the *mean excess function* is defined as

$$e(u) := E(X - u | X > u) = \frac{E((X - u)1_{(u,\infty)}(X))}{E(1_{[u,\infty)}(X))}.$$

If X has generalised Pareto distribution with $\gamma < 1$, we have

$$e(u) = \frac{\beta + \gamma u}{1 - \gamma}, \quad u \geq 0. \tag{3.9}$$

Proof.

$$\begin{aligned}
E((X-u)1_{(u,\infty)}(X)) &= \int (x-u)1_{(u,\infty)}(x)g_{\gamma,\beta}(x)dx \\
&= \int_u^\infty (x-u)g_{\gamma,\beta}(x)dx \\
&= [(x-u)(G_{\gamma,\beta}(x)-1)]_u^\infty - \int_u^\infty (G_{\gamma,\beta}(x)-1)dx \\
&= -\left[(x-u)\left(1+\frac{\gamma x}{\beta}\right)^{-\frac{1}{\gamma}}\right]_u^\infty + \int_u^\infty \left(1+\frac{\gamma x}{\beta}\right)^{-\frac{1}{\gamma}} dx \\
&= 0 + \left[\frac{\beta}{\gamma(1-\frac{1}{\gamma})}\left(1+\frac{\gamma x}{\beta}\right)^{1-\frac{1}{\gamma}}\right]_u^\infty \\
&= \frac{\beta}{1-\gamma}\left(1+\frac{\gamma u}{\beta}\right)^{1-\frac{1}{\gamma}}
\end{aligned}$$

and

$$E(1_{[u,\infty)}(X)) = P(X>u) = \overline{G}_{\gamma,\beta}(u) = \left(1+\frac{\gamma u}{\beta}\right)^{-\frac{1}{\gamma}}$$

yield the claim. □

(3.9) implies that the mean excess function is linear if X has generalised Pareto distribution with $\gamma < 1$. Is is easy to see that (3.9) holds for $u \geq u_0$ if we only assume $\overline{F}_{u_0} = \overline{G}_{\gamma,\beta}$ for the excess distribution function F_{u_0} of X.

A useful graphical tool is the *empirical mean excess function* of the observations $X_1, \ldots, X_n$, which is defined as

$$e_n(u) = \frac{1}{N_u}\sum_{k=1}^{n}(X_k - u)1_{[u,\infty)}(X_k).$$

It serves as an estimate of the unknown mean excess function of the X_i. The corresponding *mean excess plot* is the set of points

$$\{(X_{k:n}, e_n(X_{k:n})), k = 2, \ldots, n\}.$$

If this plot behaves approximately linear with positive slope beyond some threshold u_0, it seems justified to assume (3.8).

3.3.4 Estimation of the extremal cdf

Suppose that $X_1, \ldots, X_n$ are iid random variables whose excess distribution over some threshold u can be assumed to be generalised Pareto with parameters γ, β. We choose u

graphically as explained in 3.3.3 and the MLE $\hat{\gamma}, \hat{\beta}$ as in 3.3.2. Note that $\frac{N_u}{n}$ is a reasonable estimate for $\overline{F}(u)$. (3.6) and (3.8) yield

$$\begin{aligned}\overline{F}(u+x) &= \overline{F}(u)\overline{F}_u(x) \\ &\approx \frac{N_u}{n}\left(1+\frac{\gamma x}{\beta}\right)^{-\frac{1}{\gamma}} \\ &\approx \frac{N_u}{n}\left(1+\frac{\hat{\gamma} x}{\hat{\beta}}\right)^{-\frac{1}{\hat{\gamma}}}.\end{aligned}$$

Therefore,

$$\widehat{\overline{F}(u+x)} = \frac{N_u}{n}\left(1+\frac{\hat{\gamma} x}{\hat{\beta}}\right)^{-\frac{1}{\hat{\gamma}}}$$

provides a natural estimate for $\overline{F}(u+x)$, i.e. for the tails of the cdf where data is scarce.

3.3.5 Estimation of VaR

Since

$$\begin{aligned}q_p(F) &= \inf\{x \in \mathbb{R} : \overline{F}(x) \leq 1-p\} \\ &= \inf\{u+x \in \mathbb{R} : \overline{F}(u+x) \leq 1-p\}\end{aligned}$$

a natural estimator for the quantile and hence the VaR is

$$\begin{aligned}\widehat{\mathrm{VaR}_p(F)} &= \widehat{q_p(F)} \\ &:= \inf\left\{u+x \in \mathbb{R} : \widehat{\overline{F}(u+x)} \leq 1-p\right\} \\ &= u + \inf\left\{x \in \mathbb{R} : \frac{N_u}{n}\left(1+\frac{\hat{\gamma} x}{\hat{\beta}}\right)^{-\frac{1}{\hat{\gamma}}} \leq 1-p\right\} \\ &= u + \frac{\hat{\beta}}{\hat{\gamma}}\left(\left(\frac{n}{N_u}(1-p)\right)^{-\hat{\gamma}} - 1\right).\end{aligned}$$

3.3.6 Estimation of expected shortfall

We will show that

$$\widehat{\mathrm{ES}_p(F)} = \widehat{q_p(F)} + \frac{\hat{\beta} + \hat{\gamma}(\widehat{q_p(F)} - u)}{1-\hat{\gamma}}$$

is a reasonable estimator for the expected shortfall for sufficiently large u.

To this end, recall that

$$\begin{aligned}\mathrm{ES}_p(F) &= E(X|X > \mathrm{VaR}_p(F)) \\ &= \frac{E(X 1_{(q_p(F),\infty)}(X))}{E(1_{(q_p(F),\infty)}(X))}.\end{aligned}$$

Moreover, note that

$$E(Z) = \int_0^\infty P(Z > z)dz$$

for nonnegative random variables because

$$\begin{aligned}\int_0^\infty P(Z > z)dz &= \int_0^\infty E\big(1_{\{Z>z\}}\big)\, dz \\ &= E\bigg(\int_0^\infty 1_{\{Z>z\}} dz\bigg) \\ &= E(Z).\end{aligned}$$

We conclude

$$\begin{aligned}E(X1_{(q_p(F),\infty)}(X)) &= \int_0^\infty P(X1_{(q_p(F),\infty)}(X) > z)dz \\ &= q_p(F)\overline{F}(q_p(F)) + \int_{q_p(F)}^\infty \overline{F}(z)dz \\ &= q_p(F)\overline{F}(q_p(F)) + \int_{q_p(F)}^\infty \overline{F}(u)\overline{F}_u(z-u)dz.\end{aligned}$$

Moreover,

$$E(1_{(q_p(F),\infty)}(X)) = P(X > q_p(F)) = \overline{F}(q_p(F)).$$

Together, we have

$$\begin{aligned}\mathrm{ES}_p(F) &= q_p(F) + \frac{\overline{F}(u)}{\overline{F}(q_p(F))}\int_{q_p(F)}^\infty \overline{F}_u(z-u)dz \\ &= q_p(F) + \frac{1}{\overline{F}_u(q_p(F)-u)}\int_{q_p(F)}^\infty \overline{F}_u(z-u)dz.\end{aligned}$$

If we use the estimators $\widehat{\overline{F}_u(t-u)} = \overline{G}_{\hat\gamma,\hat\beta}(t-u)$ and $\widehat{q_p(F)}$ from 3.3.5, we obtain the estimator

$$\begin{aligned}\widehat{\mathrm{ES}_p(F)} &= \widehat{q_p(F)} + \frac{\int_{\widehat{q_p(F)}}^\infty \left(1+\frac{\hat\gamma(z-u)}{\hat\beta}\right)^{-\frac{1}{\hat\gamma}} dz}{\left(1+\frac{\hat\gamma(\widehat{q_p(F)}-u)}{\hat\beta}\right)^{-\frac{1}{\hat\gamma}}} \\ &= \widehat{q_p(F)} + \frac{\hat\beta + \hat\gamma(\widehat{q_p(F)}-u)}{1-\hat\gamma}\end{aligned}$$

as desired.

Remark 3.4 Summing up, appying the POT method to risk management means to

1. choose a sufficiently large threshold u based on the mean excess plot, determine N_u,
2. compute the MLE for γ, β based on the excesses $Y_1, \ldots, Y_{N_u}$,
3. use it to compute estimates for $\overline{F}(u+x)$, $\mathrm{VaR}_p(F)$, $\mathrm{ES}_p(F)$.

Chapter 4

Multivariate distributions in risk measurement

So far we have considered univariate data as e.g. a stock price. Now, we turn to multivariate data as e.g. the price vector of several stocks.

4.1 Multivariate distributions and dependence

4.1.1 Random vectors and their law

1. Let $X = (X_1, \ldots, X_d)$ be a *random vector*, i.e. a $\mathbb{R}^d$-valued random variable. Put differently, $X_1, \ldots, X_d$ are univariate random variables. The law of X is characterised by its *(joint) distribution function* $F : \mathbb{R}^d \to [0, 1]$ defined by

$$F(x) = F(x_1, \ldots, x_d) := P(X_1 \leq x_1, \ldots, X_d \leq x_d) = P(X \leq x).$$

   Its *marginal laws* are the laws of the univariate random variables X_i. The corresponding cdf's $F_i : \mathbb{R} \to [0, 1]$ are given by

$$F_i(x_i) = P(X_i \leq x_i) = F(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty).$$

   Observe that the F_i are determined by F, but not vice versa.

2. F is called *absolutely continuous* if it can be written as

$$F(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(u_1, \ldots, x_d) du_d \ldots du_1.$$

   for some function $f : \mathbb{R}^d \to \mathbb{R}_+$. This holds e.g. if $\frac{\partial^d F(x_1,\ldots,x_d)}{\partial x_1 \ldots \partial x_d}$ exists and is continuous. f is called *density* of F (resp. of X or of the law of X). If F is known, f can be obtained from F by differentiation:

$$f(x_1, \ldots, x_d) = \frac{\partial^d F(x_1, \ldots, x_d)}{\partial x_1 \ldots \partial x_d}.$$

3. The components $X_1, \ldots, X_d$ of X are stochastically independent if and only if

$$F(x_1, \ldots, x_d) = \prod_{i=1}^{d} F_i(x_i)$$

for any $x_1, \ldots, x_d$. If F has density f, this holds if and only if

$$f(x_1, \ldots, x_d) = \prod_{i=1}^{d} f_i(x_i)$$

for any $x_1, \ldots, x_d$, where $f_i : \mathbb{R} \to \mathbb{R}_+$ denotes the density of X_i, i.e. $F_i(x_i) = \int_{-\infty}^{x_i} f_i(u)du$ resp. $f_i(x_i) = F_i'(x_i)$.

4. The law of X is uniquely characterised by its *characteristic function* $\varphi_X : \mathbb{R}^d \to \mathbb{C}$, which is defined as

$$\varphi_X(t) = \varphi_X(t_1, \ldots, t_d) = E\left(\exp\left(i\sum_{j=1}^{d} t_j X_j\right)\right)$$

for $t = (t_1, \ldots, t_d) \in \mathbb{C}$.

5. The components $X_1, \ldots, X_d$ of X are stochastically independent if and only if

$$\varphi_X(t_1, \ldots, t_d) = \prod_{j=1}^{d} \varphi_{X_j}(t_j)$$

for any $t_1, \ldots, t_d \in \mathbb{R}$, where φ_{X_j} denotes the characteristic function of X_j.

4.1.2 The multivariate normal distribution

The *multivariate normal distribution* $N_d(\mu, \Sigma)$ with expectation $\mu \in \mathbb{R}^d$ and symmetric, positive definite covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ (i.e. with $x^\top \Sigma x > 0$ for any $x \neq 0$) is defined by its density

$$f(x) = \frac{1}{\sqrt{(2\pi)^d |\det \Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right), \quad x \in \mathbb{R}^d.$$

Its characteristic function equals

$$\varphi(t) = \exp\left(it^\top \mu - \frac{1}{2} t^\top \Sigma t\right), \quad t \in \mathbb{R}^d.$$

Remark 4.1 Denote by $S_n = (S_{n,1}, \ldots, S_{n,d})$ the prices of assets $1, \ldots, d$ at times $n = 0, 1, 2, \ldots$ The corresponding *logarithmic returns* $X_n = (X_{n,1}, \ldots, X_{n,d})$ are given by $X_{n,i} = \log \frac{S_{n,i}}{S_{n-1,i}}$. A common model for stock prices and foreign exchange rates is to assume that

1. $X_1, X_2, \ldots$ are iid (but not $X_{n,1}, \ldots X_{n,d}$),
2. X_n has a multivariate normal distribution, i.e. $X_n \sim N_d(\mu, \Sigma)$ for some μ, Σ.

This model is often criticised because it is inconsistent with often observed so-called *stylised facts* of financial data, namely

1. *(volatility clustering)* large absolute returns tend to cluster,
2. *(heavy tails and tail dependence)* the tails of the returns seem to be heavier and higher correlated than the components of multivariate Gaussian random vectors.

4.1.3 Comonotonicity and countermonotonicity

Comonotone random variables are basically increasing functions of each other.

Definition 4.2 Let X_1, X_2 be random variables. They are called *comonotone* if (X_1, X_2) has the same law as $(\alpha(Z), \beta(Z))$ for some random variable Z and some increasing functions $\alpha, \beta : \mathbb{R} \to \mathbb{R}$. If α is increasing and β is decreasing, X_1, X_2 are called *countermonotone* instead.

Lemma 4.3 *If X_1, X_2 have continuous cdf's F_1, F_2 and X_1, X_2 are comonotone, then X_2 is an increasing function of X_1, namely $X_2 = T(X_1)$ with $T = F_2^{\leftarrow} \circ F_1$. If X_1, X_2 are countermonotone instead of comonotone, X_2 is a decreasing function of X_1, namely $X_2 = T(X_1)$ with $T = F_2^{\leftarrow} \circ (1 - F_1)$.*

Proof. We show the statement in the comonotone case. To this end, consider $Y_1 := \alpha(Z), Y_2 := \beta(Z)$ as in the previous definition. Functions α, β are strictly increasing (outside some P^Z-null set) because F_1, F_2 are continuous. Consequently, $Y_2 = \beta(\alpha^{-1}(Y_1)) = T(Y_1)$ with $T = \beta \circ \alpha^{-1}$ and accordingly $Y_1 = \alpha(\beta^{-1}(Y_2))$. It follows that

$$\begin{aligned} F_1(x_1) &= P(\alpha(\beta^{-1}(Y_2)) \leq x_1) \\ &= P(Y_2 \leq \beta(\alpha^{-1}(x_1))) \\ &= F_2(\beta(\alpha^{-1}(x_1))) \\ &= F_2(T(x_1)), \end{aligned}$$

which implies that $T(x_1) = F_2^{-1}(F_1(x_1))$. Since $(Y_1, Y_2) \sim (X_1, X_2)$, we have that $X_2 = T(X_1)$ a.s. for $T = F_2^{-1} \circ F_1$. □

4.1.4 Moments and correlation

Let $X = (X_1, \ldots, X_d)$ be a random vector with finite second moments $E(X_i^2) < \infty$, $i = 1, \ldots, d$.

1. $E(X) := (E(X_1), \ldots, E(X_d))$ is the *expectation* of X.

2. $\mathrm{Cov}(X) := E((X-E(X))(X-E(X))^\top) = (\mathrm{Cov}(X_i, X_j))_{i,j=1,\dots,d}$ is the *covariance matrix* of X. Note that $\mathrm{Cov}(X)_{ii} = \mathrm{Var}(X_i)$ for $i = 1, \dots, d$.

3. $$\varrho_L(X_1, X_2) := \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}}$$
is the *(linear) correlation* of X_1, X_2, which is defined for random variables with nonzero variance.

4. (a) $X \mapsto E(X)$ is linear, i.e. $E(AX + b) = AE(X) + b$ for $A \in \mathbb{R}^{k\times d}, b \in \mathbb{R}^k$.
   (b) $\mathrm{Cov}(AX + b) = A\mathrm{Cov}(X)A^\top$ for $A \in \mathbb{R}^{k\times d}$.
   (c) If X_1, X_2 are independent, they are also *uncorrelated* in the sense that $\varrho_L(X_1, X_2) = 0$ or, more generally, $\mathrm{Cov}(X_1, X_2) = 0$. The converse is generally not true.
   (d) $|\varrho_L(X_1, X_2)| = 1$ holds if and only if $X_2 = a + bX_1$ with $b \neq 0$.

Example 4.4 Let X_1 be standard Gaussian and $X_2 = X_1^2$. X_1, X_2 are not independent, e.g. because $P(|X_1| > 1, |X_2| > 1) = P(|X_1| > 1) \neq P(|X_1| > 1)P(|X_2| > 1)$. However, since

$$\mathrm{Cov}(X_1, X_2) = E((X_1 - 0)(X_2 - 1)) = E(X_1X_2) = E(X_1^3) = \int_{-\infty}^{\infty} x^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0,$$

their correlation is zero, i.e. they are uncorrelated.

Remarks 4.5
1. In general $\varrho_L(T(X_1), T(X_2)) = \varrho_L(X_1, X_2)$ does *not* hold for strictly increasing functions T (except for T that are affine).

2. Suppose that X_1, X_2 are random variables with cdf's F_1, F_2. The possible correlations of X_1, X_2 belong to an interval $[\varrho_{\min}, \varrho_{\max}]$ which depends on F_1, F_2 and satisfies $0 \in [\varrho_{\min}, \varrho_{\max}] \subset [-1, 1]$. The maximal value $\varrho_{\max}$ is attained if X_1, X_2 are comonotone. The minimal value $\varrho_{\min}$, on the other hand, is attained if X_1, X_2 are countermonotone. Depending on F_1, F_2, the interval may be very small. In particular, a small correlation $\varrho_L(X_1, X_2)$ does *not* imply that X_2 cannot be an increasing function of X_1.

4.1.5 Rank correlation

Here, we discuss alternative coefficients that measure how strongly random variables X_1, X_2 are "correlated" in some sense.

Definition 4.6 *Kendall's* τ of random variables X_1, X_2 is defined as

$$\varrho_\tau(X_1, X_2) := P\Big((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) > 0\Big) - P\Big((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) < 0\Big),$$

where $(\widetilde{X}_1, \widetilde{X}_2)$ denotes a random vector, which is independent of (X_1, X_2) but has the same law.

A large $\varrho_\tau(X_1, X_2)$ indicates that if X_1 is large, then X_2 tends to be large as well. A small (i.e. large negative) $\varrho_\tau(X_1, X_2)$ indicates that if X_1 is large, then X_2 tends to be rather small.

Definition 4.7 *Spearman's* ϱ of random variables X_1, X_2 is defined as

$$\varrho_S(X_1, X_2) := \varrho_L(F_1(X_1), F_2(X_2)),$$

where F_1, F_2 denotes the cdf's of X_1, X_2.

Lemma 4.8 *Let X_1, X_2 be random variables with cdf's F_1, F_2.*

1. *$\varrho_\tau(X_1, X_2)$ and $\varrho_S(X_1, X_2)$ have values in $[-1, 1]$. For continuous cdf's all values in the interval can be attained.*

2. *Suppose that F_1, F_2 are continuous. $\varrho_\tau(X_1, X_2) = 1$ holds if and only if $\varrho_S(X_1, X_2) = 1$, which in turn holds if and only if X_1, X_2 are comonotone. Similarly, $\varrho_\tau(X_1, X_2) = -1$ holds if and only if $\varrho_S(X_1, X_2) = -1$, which in turn holds if and only if X_1, X_2 are countermonotone.*

3. *If X_1, X_2 are independent, then $\varrho_\tau(X_1, X_2) = 0$ and $\varrho_S(X_1, X_2) = 0$. The converse is not true.*

4. *If $T_1, T_2 : \mathbb{R} \to \mathbb{R}$ are strictly increasing, then $\varrho_\tau(T_1(X_1), T_2(X_2)) = \varrho_\tau(X_1, X_2)$ and $\varrho_S(T_1(X_1), T_2(X_2)) = \varrho_S(X_1, X_2)$.*

4.1.6 Estimation of linear and rank correlation

Suppose that n iid random pairs $(X_{k,1}, X_{k,2}), k = 1, \ldots, n$ are given, having the same law as (X_1, X_2).

In order to estimate the linear correlation $\varrho_L(X_1, X_2)$, recall that the sample variance

$$\widehat{\mathrm{Var}}(X_1) = \frac{1}{n-1} \sum_{k=1}^{n} (X_{k,1} - \overline{X}_1)^2$$

with $\overline{X}_1 = \frac{1}{n} \sum_{k=1}^{n} X_{k,1}$ is a reasonable estimator for $\mathrm{Var}(X_1)$ and accordingly for X_2. Similarly,

$$\widehat{\mathrm{Cov}}(X_1, X_2) = \frac{1}{n-1} \sum_{k=1}^{n} (X_{k,1} - \overline{X}_1)(X_{k,2} - \overline{X}_2)$$

is the standard estimator for $\mathrm{Cov}(X_1, X_2)$. This leads to the natural estimator

$$\hat{\varrho}_L(X_1, X_2) = \frac{\widehat{\mathrm{Cov}}(X_1, X_2)}{\sqrt{\widehat{\mathrm{Var}}(X_1)\widehat{\mathrm{Var}}(X_2)}}$$

for $\varrho_L(X_1, X_2)$.

In order to estimate Kendall's τ, observe that

$$\begin{aligned}\varrho_\tau(X_1,X_2) &= P\Big((X_1-\widetilde{X}_1)(X_2-\widetilde{X}_2)>0\Big)-P\Big((X_1-\widetilde{X}_1)(X_2-\widetilde{X}_2)<0\Big)\\ &= E\Big(\mathrm{sgn}((X_1-\widetilde{X}_1)(X_2-\widetilde{X}_2))\Big).\end{aligned}$$

The idea now is to replace the expectation by the empirical mean. This leads to the estimator

$$\begin{aligned}\hat{\varrho}_\tau(X_1,X_2) &= \frac{1}{\binom{n}{2}}\sum_{1\le k<\ell\le n}\mathrm{sgn}((X_{k,1}-X_{\ell,1})(X_{k,2}-X_{\ell,2}))\\ &= \frac{1}{\binom{n}{2}}\sum_{k=1}^{n-1}\sum_{\ell=k+1}^{n}\mathrm{sgn}((X_{k,1}-X_{\ell,1})(X_{k,2}-X_{\ell,2})).\end{aligned}\tag{4.1}$$

Spearman's ϱ can be estimated by

$$\hat{\varrho}_S(X_1,X_2)=\frac{12}{n(n^2-1)}\sum_{k=1}^{n}\left(\mathrm{rank}(X_{k,1})-\frac{n+1}{2}\right)\left(\mathrm{rank}(X_{k,2})-\frac{n+1}{2}\right)\tag{4.2}$$

where $\mathrm{rank}(X_{k,1})$ denotes the rank of $X_{k,1}$ in $X_{1,1},\dots,X_{n,1}$, e.g. $\mathrm{rank}(X_{k,1})=3$ if it is the third smallest observation. In order to justify it, recall that $F_1(X_1)$ and $F_2(X_2)$ have uniform law on $[0,1]$ for continuous distributions because

$$P(F_1(X_1)\le t)=P(X_1\le F_1^{\leftarrow}(t))=F_1(F_1^{\leftarrow}(t))=t$$

and likewise for $F_2(X_2)$. This implies

$$\begin{aligned}\varrho_S(X_1,X_2) &= \varrho_L(F_1(X_1),F_2(X_2))\\ &= \frac{E\big((F_1(X_1)-E(F_1(X_1)))(F_2(X_2)-E(F_2(X_2)))\big)}{\sqrt{\mathrm{Var}(F_1(X_1))\mathrm{Var}(F_2(X_2))}}\\ &= 12E\big((F_1(X_1)-E(F_1(X_1)))(F_2(X_2)-E(F_2(X_2)))\big).\end{aligned}$$

Let us now replace F_1,F_2 by their empirical counterparts $\hat{F}_1,\hat{F}_2$ and expectations by the empirical mean. We have

$$\hat{F}_1(X_{k,1})=\frac{1}{n}\mathrm{rank}(X_{k,1})$$

and hence

$$\frac{1}{n}\sum_{k=1}^{n}\hat{F}_1(X_{k,1})=\frac{1}{n}\sum_{k=1}^{n}\frac{1}{n}\mathrm{rank}(X_{k,1})=\frac{1}{n^2}\sum_{k=1}^{n}k=\frac{n+1}{2n}$$

as an estimate for $E(F_1(X_1))$ and accordingly for $E(F_2(X_2))$. This leads to the estimator

$$\frac{12}{n}\sum_{k=1}^{n}\left(\frac{1}{n}\mathrm{rank}(X_{k,1})-\frac{n+1}{2n}\right)\left(\frac{1}{n}\mathrm{rank}(X_{k,2})-\frac{n+1}{2n}\right)=\frac{n^2-1}{n^2}\hat{\varrho}_S(X_1,X_2),$$

which is almost the one in (4.2).

4.1.7 Tail dependence

Let X_1, X_2 be random variables with cdf's F_1, F_2. The *upper tail dependence coefficient* is defined as

$$\lambda_u(X_1, X_2) = \lim_{u \uparrow 1} P(X_2 > F_2^{\leftarrow}(u) | X_1 > F_1^{\leftarrow}(u))$$

if the limit exists. Similarly, the *lower tail dependence coefficient* is defined as

$$\lambda_\ell(X_1, X_2) = \lim_{u \downarrow 0} P(X_2 < F_2^{\leftarrow}(u) | X_1 < F_1^{\leftarrow}(u))$$

if the limit exists.

If $\lambda_u(X_1, X_2) > 0$ (resp. $\lambda_\ell(X_1, X_2) > 0$), one says that X_1, X_2 have upper (resp. lower) tail dependence.

Remarks 4.9

1. $\lambda_u(X_1, X_2)$ and $\lambda_\ell(X_1, X_2)$ are in $[0, 1]$ if they exist.

2. In order to illustrate tail dependence, consider $u = 0.99$. Then $P(X_2 > F_2^{\leftarrow}(u) | X_1 > F_1^{\leftarrow}(u))$ stands for the probability that X_2 represents a large loss (i.e. appearing once in 100 observations) given that X_1 represents a large loss (i.e. appearing once in 100 observations).

3. If F_1 is continuous, we have

$$P(X_1 > F_1^{\leftarrow}(u)) = 1 - P(X_1 \leq F_1^{\leftarrow}(u)) = 1 - F_1(F_1^{\leftarrow}(u)) = 1 - u$$

and hence

$$P(X_2 > F_2^{\leftarrow}(u) | X_1 > F_1^{\leftarrow}(u)) = \frac{P(X_2 > F_2^{\leftarrow}(u), X_1 > F_1^{\leftarrow}(u))}{1 - u}.$$

Therefore, $\lambda_u(X_1, X_2)$ and likewise $\lambda_\ell(X_1, X_2)$ are symmetric in X_1, X_2 if their cdf's F_1, F_2 are continuous.

4.2 Elliptical distributions

Elliptical distributions generalise the multivariate normal distribution while keeping some of its tractability.

4.2.1 The multivariate normal distribution

Let us recall some properties of the multivariate normal distribution.

1. By definition, random vector $X = (X_1, \ldots, X_d)$ is *multivariate normal* if and only if the univariate random variable $a^\top X = \sum_{i=1}^d a_i X_i$ is normally distributed for any $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$.

2. The notation $X \sim N_d(\mu, \Sigma)$ indicates that X is an $\mathbb{R}^d$-valued random vector with $E(X) = \mu \in \mathbb{R}^d$, $\mathrm{Cov}(X) = \Sigma \in \mathbb{R}^{d\times d}$.

3. $X \sim N_d(\mu, \Sigma)$ holds if and only if the characteristic function of X is of the form

$$\varphi_X(u) = E\big(\exp(iu^\top X)\big) = \exp\left(iu^\top\mu - \frac{1}{2}u^\top\Sigma u\right), \quad u \in \mathbb{R}^d.$$

4. Let $\mu \in \mathbb{R}^d$ be a vector and $\Sigma \in \mathbb{R}^{d\times d}$ a symmetric, positive definite matrix. Then $X \sim N_d(\mu, \Sigma)$ holds if and only if

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^d|\det\Sigma|}}\exp\left(\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right)$$

holds for the density function.

5. Let $X \sim N_d(\mu, \Sigma)$.

   (a) (Linear transformations) Then $BX + b \sim N_k(B\mu + b, B\Sigma B^\top)$ for $b \in \mathbb{R}^k$, $B \in \mathbb{R}^{k\times d}$.

   (b) (Marginal laws) Let $X_{(1)} = (X_1, \ldots, X_k)$ and $X_{(2)} = (X_{k+1}, \ldots, X_d)$, which implies $X = (X_{(1)}, X_{(2)})$. Moreover, write

$$\mu = (X_{(1)}, X_{(2)}), \quad \Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix}.$$

   Then $X_{(1)} \sim N_k(\mu_{(1)}, \Sigma_{(11)})$ and $X_{(2)} \sim N_{d-k}(\mu_{(2)}, \Sigma_{(22)})$.

   (c) (Conditional laws) Suppose that $\det\Sigma \neq 0$. Then the conditional law of $X_{(2)}$ given $X_{(1)} = x_1$ is

$$P^{X_{(2)}|X_{(1)}=x_1} = N_{d-k}(\mu_{2,1}, \Sigma_{22,1}),$$

   where $\mu_{2,1} = \mu_{(2)} + \Sigma_{(21)}\Sigma_{(11)}^{-1}(x_1 - \mu_{(1)})$ and $\Sigma_{22,1} = \Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)}$.

   (d) (Quadratic forms) Suppose that $\det\Sigma \neq 0$. Then $D^2 := (X-\mu)^\top\Sigma^{-1}(X-\mu)$ has a chi-square distribution with d degrees of freedom. The random variable D is called *Mahalanobis distance*.

   (e) (Convolution) If $X \sim N_d(\mu, \Sigma)$ and $Y \sim N_d(\widetilde{\mu}, \widetilde{\Sigma})$ are independent, we have $X + Y \sim N_d(\mu + \widetilde{\mu}, \Sigma + \widetilde{\Sigma})$.

   (f) (Representation based on standard normal vectors) X has the representation $X = \mu + AZ$ with $Z \sim N_k(0, 1_k)$ for some $k \leq d$ and some $A \in \mathbb{R}^{d\times k}$ with $AA^\top = \Sigma$. If $\det\Sigma \neq 0$, then $k = d$.

A first generalisation concerns mixtures of normals.

Definition 4.10 An $\mathbb{R}^d$-valued random vector X is distributed according to a *multivariate normal mixture* if it is of the form $X = \mu + WAZ$, where $Z \sim N_k(0, 1_k)$, $A \in \mathbb{R}^{d\times k}$, $\mu \in \mathbb{R}^d$ and $W \geq 0$ is a random variable which is independent of Z.

Remark 4.11 In the situation of the previous definition, the conditional law of X given $W = w$ equals $P^{X|W=w} = N_d(\mu, w^2\Sigma)$, where $\Sigma = AA^\top$.

Example 4.12 Let S be chi-square distributed with ν degrees of freedom and set $W = \sqrt{\frac{\nu}{S}}$. Then $X \sim t_d(\nu, \mu, \Sigma)$, i.e. X has a multivariate t-distribution with ν degrees of freedom. Note that in this case Σ is not the covariance matrix of X. Instead, we have $\mathrm{Cov}(X) = \frac{\nu}{\nu-2}\Sigma$ because $E(W^2) = \frac{\nu}{\nu-2}$ for $\nu > 2$.

4.2.2 Spherical distributions

Spherical distributions put equal mass in any direction from the origin.

Definition 4.13 A random vector $X = (X_1, \ldots, X_d)$ is *spherically distributed* if its characteristic function is of the form

$$\varphi_X(u) = \psi(u^\top u) = \psi(u_1^2 + \cdots + u_d^2), \quad u \in \mathbb{R}^d$$

for some function $\psi : \mathbb{R} \to \mathbb{C}$. In this case we write $X \sim S_d(\psi)$.

Theorem 4.14 *The following statements are equivalent.*

1. *X has a spherical distribution.*
2. *$a^\top X$ has the same law as $|a|X_1$ for any $a \in \mathbb{R}^d$, where $|a| = \sqrt{a_1^2 + \cdots + a_d^2}$.*
3. *$X = RS$, where S is uniformly distributed on the sphere $S^{d-1} = \{x \in \mathbb{R}^d : |x| = 1\}$ and $R \geq 0$ is a random variable that is independent of S.*

Example 4.15 Let $X \sim N_d(0, 1_d)$. Then $X \sim S_d(\psi)$ for $\psi(x) = \exp(-\frac{x}{2})$ because $\varphi_X(u) = \exp(iu^\top 0 - \frac{1}{2}u^\top 1_d u) = \exp(-\frac{1}{2}u^\top u) = \psi(u^\top u)$. In the representation $X = RS$, we have that $\sum_{i=1}^d X_i^2 = |X|^2 = R^2$ because $|S| = 1$. This implies that R^2 is chi-square distributed with d degrees of freedom.

In order to simulate spherically distributed random vectors, we use the representation $X = RS$ and proceed in the following steps:

1. First draw a realisation s of S, i.e. from a uniform distribution on the sphere. To this end, note that $\frac{Z}{|Z|}$ has the same law as S for standard normal random variables $Z_1, \ldots, Z_d$ and $Z = (Z_1, \ldots, Z_d)$. Hence, we can choose $s = \frac{(z_1,\ldots,z_d)}{\sqrt{z_1^2+\cdots+z_d^2}}$, where $z_1, \ldots, z_d$ are independent realisations from the standard normal distribution.
2. Draw a realisation r from the law of R.
3. Then $x = rs$ is a realisation from the law of X.

4.2.3 Elliptical distributions

Elliptical distributions resemble multivariate normal laws in certain respects.

Definition 4.16 An $\mathbb{R}^d$-valued random vector X is *elliptically distributed* if $X = \mu + AY$, where $Y \sim S_k(\psi)$, $A \in \mathbb{R}^{d\times k}$, $\mu \in \mathbb{R}^d$. We write $X \sim E_d(\mu, \Sigma, \psi)$. μ is called *location parameter*, $\Sigma := AA^\top$ *dispersion matrix*, and ψ *characteristic generator*.

Remarks 4.17 1. The corresponding characteristic function equals

$$\begin{aligned}\varphi_X(u) &= E(\exp(iu^\top X)) \\ &= E(\exp(iu^\top(\mu + AY))) \\ &= \exp(iu^\top\mu)E(\exp(i(A^\top u)^\top Y)) \\ &= \exp(iu^\top\mu)\psi(u^\top AA^\top u) \\ &= \exp(iu^\top\mu)\psi(u^\top\Sigma u).\end{aligned}$$

2. $E(X) = \mu$ if the expectation exists.

3. $\mathrm{Cov}(X) = c\Sigma$ for some $c > 0$ if X has finite second moments.

4. $X - \mu$ is symmetric in the sense that $X - \mu$ and $\mu - X$ have the same law.

5. If $\Sigma = AA^\top$ with invertible $A \in \mathbb{R}^{d\times d}$, then $X \sim E_d(\mu, \Sigma, \psi)$ holds if and only if $A^{-1}(X - \mu) \sim S_d(\psi)$. In general $X \sim E_d(\mu, \Sigma, \psi)$ holds if and only if $X = \mu + RAS$ for some S which is uniform on the sphere, $A \in \mathbb{R}^{d\times k}$ with $AA^\top = \Sigma$, and some nonnegative random variable $R \geq 0$ that is independent of S (and hence $RS \sim S_k(\psi)$).

6. In order to simulate elliptical random vectors, we use the representation $X = \mu+RAS$ and proceed in the following steps:

   (a) Simulate a realisation s of S as explained at the end of Section 4.2.2.

   (b) Simulate a realisation r of R.

   (c) Obtain a realisation $x = \mu + rAs$ of X.

Normal distributions and mixtures of normals are elliptical:

Example 4.18 $X \sim N_d(\mu, \Sigma)$ is elliptic, more specifically $X \sim E_d(\mu, \Sigma, \psi)$ with $\psi(x) = \exp(-\frac{x}{2})$. Indeed, we have $X = \mu + AZ$ with $AA^\top = \Sigma$ and $Z \sim N_k(0, 1_k)$. Note that $Z \sim S_k(\psi)$ by Example 4.15.

Example 4.19 If X has a normal mixture distribution, then it is elliptical. Indeed, let $X = \mu + WAZ$ with $W \geq 0$, $Z \sim N_k(0, 1_k)$, W, Z being independent, and $A \in \mathbb{R}^{d\times k}$, $\mu \in \mathbb{R}^d$. By Example 4.15, we have that $Z = VS$, where V^2 is chi-square distributed with k degrees

of freedom, S is uniform on the sphere, and V, S are independent. Consequently, $X = \mu + VWAS = \mu + RAS$. This implies $X \sim E_d(\mu, \Sigma, \psi)$ with $\Sigma = AA^\top$ and $RS = S_k(\psi)$.

As an example, consider $W^2 \sim \frac{\nu}{\widetilde{S}}$ with some chi-square distributed $\widetilde{S}$ with ν degrees of freedom as in Example 4.12. Then $X \sim t_d(\nu, \mu, \Sigma)$. Moreover, $\frac{R^2}{k} = \frac{V^2 W^2}{k}$ has an $F(k, \nu)$ distribution because $V^2 \sim \chi^2_k$, $\frac{\nu}{W^2} \sim \chi^2_\nu$, and V, W are independent.

Elliptical distributions have the following properties which resemble those of the normal distribution.

Theorem 4.20 *Let $X \sim E_d(\mu, \Sigma, \psi)$.*

1. *(Linear transformations) $BX + b \sim E_k(B\mu + b, B\Sigma B^\top, \psi)$ for $b \in \mathbb{R}^k$, $B \in \mathbb{R}^{k \times d}$.*
2. *(Marginals) Let $X_{(1)} = (X_1, \ldots, X_k)$ and $X_{(2)} = (X_{k+1}, \ldots, X_d)$, which implies $X = (X_{(1)}, X_{(2)})$. Moreover, write*

$$\mu = (\mu_{(1)}, \mu_{(2)}), \quad \Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix}.$$

   Then $X_{(1)} \sim E_k(\mu_{(1)}, \Sigma_{(11)}, \psi)$ and $X_{(2)} \sim E_{d-k}(\mu_{(2)}, \Sigma_{(22)}, \psi)$.
3. *(Conditional distributions) Suppose that $\det \Sigma \neq 0$. Then the conditional law of $X_{(2)}$ given $X_{(1)} = x_1$ is*

$$P^{X_{(2)} | X_{(1)} = x_1} = E_{d-k}(\mu_{2,1}, \Sigma_{22,1}, \widetilde{\psi}),$$

   where $\mu_{2,1} = \mu_{(2)} + \Sigma_{(21)}\Sigma^{-1}_{(11)}(x_1 - \mu_{(1)})$ and $\Sigma_{22,1} = \Sigma_{(22)} - \Sigma_{(21)}\Sigma^{-1}_{(11)}\Sigma_{(12)}$, and $\widetilde{\psi}$ generally differs from ψ.
4. *(Quadratic forms) If $\det \Sigma \neq 0$, then $D^2 = (X - \mu)^\top \Sigma^{-1}(X - \mu)$ has the same law as R^2.*
5. *(Convolution) If $X \sim E_d(\mu, \Sigma, \psi)$ and $Y \sim E_d(\widetilde{\mu}, \Sigma, \widetilde{\psi})$ are independent, we have $X + Y \sim E_d(\mu + \widetilde{\mu}, \Sigma, \bar{\psi})$ with $\bar{\psi}(x) = \psi(x)\widetilde{\psi}(x)$.*

Remarks 4.21

1. In general, $X \sim E_d(0, 1_d, \psi)$ does not imply independence of $X_1, \ldots, X_d$ unless X is multivariate normal.
2. For elliptical distributions the VaR is subadditive in the sense that $\mathrm{VaR}_p(w_1^\top X + w_2^\top X) \leq \mathrm{VaR}_p(w_1^\top X) + \mathrm{VaR}(w_2^\top X)$ for $w_1, w_2 \in \mathbb{R}^d$ and $X \sim E_d(\mu, \Sigma, \psi)$.

4.3 Copulas

Copulas are a tool allowing to characterize the dependence of components of a random vector independently of the marginal laws. Moreover, they can be used to construct random vectors with a nonstandard dependence structure.

4.3.1 Sklar's theorem

Definition 4.22 A d-dimensional *copula* is the cumulative distribution function (defined on the set $[0,1]^d$) of a random vector $(U_1, \ldots, U_d)$ whose components $U_1, \ldots, U_d$ are uniformly distributed on $[0,1]$.

Remark 4.23 Recall the following properties of cumulative distribution functions G of real-valued random variables:

1. If random variable U is uniform on $[0,1]$, we have $P(G^{\leftarrow}(U) \leq x) = G(x)$ for $x \in \mathbb{R}$, i.e. $G^{\leftarrow}(U)$ has cdf G.
2. If random variable Y has cdf G and G is continuous, then $G(Y)$ is uniform on $[0,1]$.
3. If G is continuous, then $G(G^{\leftarrow}(u)) = u$ for $u \in (0,1)$.

We want to characterize the joint law of a random vector X uniquely by its marginal laws (represented by the cdf's $F_1, \ldots, F_n$ of $X_1, \ldots, X_n$) and its dependence structure (represented in terms of some copula C). Such a characterization is provided by the following

Theorem 4.24 (Sklar) *1. Let F denote a cdf on $\mathbb{R}^d$ with marginal cdf's $F_1, \ldots, F_d$. Then there exists a copula C satisfying*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) \tag{4.3}$$

for all $x_1, \ldots, x_d \in [-\infty, \infty]$. C is unique if $F_1, \ldots, F_d$ are continuous.

2. Let C denote a copula and $F_1, \ldots, F_d$ cdf's on $\mathbb{R}$. Then (4.3) defines a cdf on $\mathbb{R}^d$ with marginals $F_1, \ldots, F_d$ and copula C.

If $F_1, \ldots, F_d$ are continuous, the copula C in Sklar's theorem 4.24 is given by $C(u_1, \ldots, u_d) = F(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d))$.

Definition 4.25 C in Theorem 4.24 is called *copula of F*. If C is the cdf of some random vector X, we call C *copula of X*.

Lemma 4.26 *Let $X = (X_1, \ldots, X_d)$ denote a random vector with continuous cdf and copula C. Let $T_1, \ldots, T_d$ be strictly increasing functions on $\mathbb{R}$. Then C is copula of the random vector $(T_1(X_1), \ldots, T_d(X_d))$ as well.*

Example 4.27 1. Let X be multivariate normal with law $N_d(0, R)$, where R is a *correlation matrix*, i.e. a covariance matrix whose diagonal entries are all 1. The copula C_R^{Ga} of X is called *Gaussian copula* for R. It is given by

$$C_R^{\mathrm{Ga}}(u_1, \ldots, u_d) = \Phi_R(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_1)),$$

where Φ_R is the cdf of X and Φ denotes the cdf of $N(0,1)$ as usual.

2. Let Y be multivariate normal with law $N_d(\mu, \Sigma)$ where Σ is invertible: The covariance matrix can be written as $\Sigma = DRD$ with some diagonal matrix $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_d)$, $\sigma_i > 0$. Then C_R^{Ga} is the copula of Y. Indeed, we have $Y_i = \mu_i + \sigma_i X_i$ for a random vector $X = (X_1, \ldots, X_d) \sim N_d(0, R)$ as in 1. By Lemma 4.26 X and Y have the same copula.

4.3.2 Correlation coefficients and tail dependence

The dependence measures of Section 4.1.5 can be expressed in terms of copulas:

Theorem 4.28 *Let $X = (X_1, X_2)$ denote a random vector with copula C and continuous marginals.*

1. *We have*

$$\begin{aligned} \varrho_\tau(X_1, X_2) &= 4 \int_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1 \\ &= 4 \int_{[0,1]^2} C(u_1, u_2) \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} du_1 du_2 - 1 \\ &= 4E(C(U_1, U_2)) - 1 \end{aligned}$$

   if (U_1, U_2) denotes a random vector with cdf C.

2. *We have*

$$\begin{aligned} \varrho_S(X_1, X_2) &= 12 \int_{[0,1]^2} u_1 u_2 dC(u_1, u_2) - 3 \\ &= 12 \int_{[0,1]^2} C(u_1, u_2) du_1 du_2 - 3 \\ &= 12E(U_1 U_2) - 3 \end{aligned}$$

   if (U_1, U_2) denotes a random vector with cdf C.

3. *If the limits exist, we have*

$$\begin{aligned} \lambda_u(X_1, X_2) &= \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u}, \\ \lambda_\ell(X_1, X_2) &= \lim_{u \downarrow 0} \frac{C(u, u)}{u}. \end{aligned}$$

We consider two examples that will be revisited in Example 4.42.

Example 4.29 The *Gumbel copula* with parameter $\vartheta \geq 1$ is defined as

$$C_\vartheta(u_1, u_2) = \exp\Big(-\big((-\log u_1)^\vartheta + (-\log u_2)^\vartheta \big)^{\frac{1}{\vartheta}} \Big).$$

From

$$\frac{1-2u+C_\vartheta(u,u)}{1-u} = \frac{1-2u+\exp(2^{1/\vartheta}\log u)}{1-u} = \frac{1-2u+u^{(2^{1/\vartheta})}}{1-u}$$

and hence

$$\lim_{u\to 1}\frac{1-2u+C_\vartheta(u,u)}{1-u} \overset{\text{l'Hospital}}{=} \lim_{u\to 1}\left(2-2^{1/\vartheta}u^{(2^{1/\vartheta}-1)}\right) = 2-2^{1/\vartheta},$$

it follows that the upper tail dependence coefficient is $\lambda_u = 2-2^{1/\vartheta}$ (which is large if ϑ is large). The lower tail dependence coefficient, however, equals $\lambda_\ell = 0$ because

$$\lim_{u\to 0}\frac{C_\vartheta(u,u)}{u} = \lim_{u\to 0}\frac{u^{(2^{1/\vartheta})}}{u} = \lim_{u\to 0} 2^{1/\vartheta}u^{(2^{1/\vartheta}-1)} = 0$$

by l'Hospital's rule.

Example 4.30 The *Clayton copula* with parameter $\vartheta > 0$ is defined as

$$C_\vartheta(u_1,u_2) = \left(u_1^{-\vartheta}+u_2^{-\vartheta}-1\right)^{-1/\vartheta}.$$

From

$$\begin{aligned}
\lim_{u\to 0}\frac{C_\vartheta(u,u)}{u} &= \lim_{u\to 0}\frac{\left(2u^{-\vartheta}-1\right)^{-1/\vartheta}}{u} \\
&\overset{\text{l'Hospital}}{=} \lim_{u\to 0} -\frac{1}{\vartheta}(2u^{-\vartheta}-1)^{-\frac{1}{\vartheta}-1}(-2\vartheta u^{-\vartheta-1}) \\
&= \lim_{u\to 0} 2(2-u^\vartheta)^{-\frac{1}{\vartheta}-1} \\
&= 2^{-1/\vartheta}
\end{aligned}$$

it follows that the lower tail dependence coefficient is $\lambda_\ell = 2^{-1/\vartheta}$ (which is large if ϑ is large). The upper tail dependence coefficient, on the other hand, equals $\lambda_u = 0$ because

$$\lim_{u\to 1}\frac{1-2u+C_\vartheta(u,u)}{1-u} = \lim_{u\to 1}\left(2-2(2-u^\vartheta)^{-\frac{1}{\vartheta}-1}\right) = 0$$

by l'Hospital's rule.

4.3.3 Elliptical copulas

Elliptical distributions give rise to the family of so-called elliptical copulas.

Definition 4.31 Let X be elliptically distributed according to $E_d(\mu,\Sigma,\psi)$ with cdf F and continuous marginals $F_1,\ldots,F_d$. The corresponding copula

$$C(u) = F\left(F_1^{\leftarrow}(u_1),\ldots,F_d^{\leftarrow}(u_d)\right)$$

is called an *elliptical copula*.

Examples of elliptical copulas are the Gaussian copula C_R^{Ga} with correlation matrix R and the t-copula $C_{\nu,R}^t$ with parameter ν and correlation matrix R, which is introduced in Example 4.33.

The Gaussian copula is inadequate for modelling random vectors with pronounced tail dependence as the follwing result shows.

Lemma 4.32 *1. If $X = (X_1, X_2)$ is a Gaussian random vector whose components are linearly independent, then $\lambda_u(X_1, X_2) = \lambda_\ell(X_1, X_2) = 0$.*

2. Let $X = (X_1, X_2)$ be a random vector with continuous marginals and Gaussian copula. Unless X_2 is a increasing or decreasing function of X_1, we have $\lambda_u(X_1, X_2) = \lambda_\ell(X_1, X_2) = 0$.

As an example of a copula allowing for tail dependence we consider the t-copula.

Example 4.33 Let X be a $t_d(\nu, \mu, \Sigma)$-distributed random vector, i.e. $X = \mu + \frac{\sqrt{\nu}}{\sqrt{S}} AZ$ with $\mu \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$, $\Sigma = AA^\top$ and independent S, Z such that random variable S has law χ^2_ν and Z is $\mathbb{R}^d$-valued and standard Gaussian.

1. We have $E(X) = \mu$ if $\nu > 1$ and $\mathrm{Cov}(X) = \frac{\nu}{\nu-2}\Sigma$ if $\nu > 2$.

2. The copula of X depends only on ν and the *correlation matrix* R defined as

$$R_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}, \quad i, j = 1, \ldots, d.$$

   This *t-copula* corresponding to ν, R is denoted as $C_{\nu,R}^t$.

3. Since $C_{\nu,R}^t$ is the copula of $t_d(\nu, 0, R)$, it is elliptical and satisfies

$$C_{\nu,R}^t(u_1, \ldots, u_d) = t_{\nu,R}^d(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)),$$

   where $t_{\nu,R}^d$ is the multivariate cdf of $X \sim t_d(\nu, 0, R)$ and t_ν are the corresponding marginal cdf's, i.e. the cdf of $X_1 \sim t_1(\nu, 0, 1)$.

The following result shows that the t-copula allows for tail dependence. It is large for small ν and large correlations. Moreover, it is symmetric in the sense that $\lambda_u = \lambda_\ell$.

Lemma 4.34 *1. The $X = (X_1, X_2)$ be t-distributed with ν degrees of freedom and correlation matrix R. Then*

$$\lambda_u(X_1, X_2) = \lambda_\ell(X_1, X_2) = 2\overline{t_{\nu+1}}\left(\frac{\sqrt{\nu+1}\sqrt{1-R_{12}}}{\sqrt{1+R_{12}}}\right). \tag{4.4}$$

with $\overline{t_{\nu+1}} = 1 - t_{\nu+1}$.

2. If $X = (X_1, X_2)$ is a random vector with continuous marginals and t-copula with parameters ν and R, then (4.4) holds.

The follwing result concerns the correlation coefficients of elliptical copulas

Lemma 4.35 *1. Let $X = (X_1, X_2)$ be elliptically distributed according to $E_2(\mu, \Sigma, \psi)$ and with continuous marginals. Then Kendall's τ equals*

$$\varrho_\tau(X_1, X_2) = \frac{2}{\pi} \arcsin R_{12}, \text{ where } R_{12} = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}. \tag{4.5}$$

2. *If $X = (X_1, X_2)$ is a random vector with continuous marginals and the same copula as $E_2(\mu, \Sigma, \psi)$, then (4.5) holds.*

3. *For the Gaussian or G_R^{Ga}-copula, Spearman's ϱ equals $\varrho_S = \frac{6}{\pi} \arcsin(R_{12}/2)$.*

As noted above, copulas can be used for simulation purposes. By Sklar's theorem and in particular (4.3), simulation of a random vector can be reduced to simulating its copula. In the following we consider Gaussian and t-copulas.

Example 4.36 (Simulation of Gaussian copulas G_R^{Ga}) Recall that

- C_R^{Ga} is the copula of a random vector $X = (X_1, \ldots, X_d)$ with law $N_d(0, R)$, where R denotes a correlation matrix,
- X has marginal cdf Φ, i.e. all components $X_1, \ldots, X_d$ are standard normal,
- if $Z_1, \ldots, Z_d$ are independent real-valued standard normal random variables, then AZ has law $N_d(0, R)$ for $Z = (Z_1, \ldots, Z_d)$ and A with $AA^\top = R$,
- the *Cholesky decomposition* of R is $R = AA^\top$ for some lower triangular matrix A. It is implemented in many software packages.

This suggests the following algorithm for simulating random vectors with cdf C_R^{Ga}:

1. Determine the Cholesky decomposition of R, i.e. A with $R = AA^\top$.
2. Simulate d independent realisations $z_1, \ldots, z_d$ of a standard normal random variable.
3. Set $(x_1, \ldots, x_d)^\top := A(z_1, \ldots, z_d)^\top$.
4. Set $u_k := \Phi(x_k)$, $k = 1, \ldots, d$, where Φ is the cdf of $N(0, 1)$ as usual.
5. Now $u = (u_1, \ldots, u_d)$ is a realisation of a random vector with cdf C_R^{Ga}.

Example 4.37 (Simulation of t-copulas $C_{\nu,R}^t$) Recall that

- $C_{\nu,R}^t$ is the copula of a random vector X with law $t_d(\nu, 0, R)$, where R denotes a correlation matrix,
- the components of X have cdf t_ν, which denotes the cdf of $t_1(\nu, 0, 1)$,

- if $Z_1, \dots, Z_d$ are independent real-valued standard normal random variables and S is an independent random variable with law χ^2_ν, then $\sqrt{\frac{\nu}{S}}AZ$ has law $t_d(\nu, 0, R)$ for $Z = (Z_1, \dots, Z_d)$ and A with $AA^\top = R$.

This suggests the following algorithm for simulating random vectors with cdf $C^t_{\nu,R}$:

1. Determine the Cholesky decomposition of R, i.e. A with $R = AA^\top$.
2. Simulate d independent realisations $z_1, \dots, z_d$ of a standard normal random variable.
3. Simulate an independent realisation s of a χ^2_ν-distributed random variable.
4. Set $(x_1, \dots, x_d)^\top := \sqrt{\frac{\nu}{S}}A(z_1, \dots, z_d)^\top$.
5. Set $u_k := t_\nu(x_k)$, $k = 1, \dots, d$.
6. Now $u = (u_1, \dots, u_d)$ is a realisation of a random vector with cdf $C^t_{\nu,R}$.

4.3.4 Archimedean copulas

Unfortunately, closed-form expressions are typically not available for elliptical copulas. Moreover, they feature certain symmetries (e.g. $\lambda_u = \lambda_\ell$) which may not be present in real data. As an alternative we discuss Archimedean copulas which generalize the Gumbel and Clayton copulas from Examples 4.29 and 4.30. However, even if X and $-X$ may have different copulas in the Archimedean case, we have instead symmetry in the components X_i. Put differently, the copula of of X_i, X_j is identical for all pairs i, j with $i \neq j$.

Definition 4.38 A function $g : \mathbb{R}_+ \to \mathbb{R}_+$ is called *completely monotone* if it is continuous and $(-1)^k g^{(k)}(t) \geq 0$ for $k = 0, 1, 2, \dots$ and any $t > 0$.

Example 4.39 $g(t) = \exp(-t^{1/\vartheta})$ for $\vartheta \geq 1$ and $g(t) = (1 + t)^{-1/\vartheta}$ for $\vartheta > 0$ are completely monotone functions.

Theorem 4.40 *Let $\varphi : [0, 1] \to [0, \infty]$ be a continuous, strictly decreasing function such that $\varphi(0) = \infty$, $\varphi(1) = 0$, and $\varphi^{-1} : \mathbb{R}_+ \to \mathbb{R}_+$ is completely monotone. Moreover, fix $d \geq 2$. Then $C : [0, 1]^d \to [0, 1]$ defined by $C(u) = \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_d))$ is a copula. For $d = 2$, complete monotonicity can be replaced by the weaker requirement of convexity.*

Definition 4.41 C in Theorem 4.40 is called *Archimedean copula* with generator φ.

Example 4.42 1. The family of *Gumbel copulas* with parameter $\vartheta \geq 1$ is obtained for $\varphi(u) = (-\log u)^\vartheta$. We obtain

$$C^{\mathrm{Gu}}_\vartheta(u_1, \dots, u_d) := \exp\left(-\left(\sum_{i=1}^d (-\log u_i)^\vartheta \right)^{1/\vartheta} \right),$$

cf. Example 4.29 for $d = 2$. They have the following properties:

(a) upper tail dependence, cf. Example 4.29,

(b) $C_1^{\mathrm{Gu}}(u_1, \ldots, u_d) = \prod_{i=1}^d u_i$, which is the copula for independent random variables,

(c) $\lim_{\vartheta \to \infty} C_\vartheta^{\mathrm{Gu}}(u_1, \ldots, u_d) = \min\{u_1, \ldots, u_d\}$, which is the copula for comonotone random variables,

(d) symmetry in $u_1, \ldots, u_d$.

2. The family of *Clayton copulas* with parameter $\vartheta > 0$ is obtained for $\varphi(u) = u^{-\vartheta} - 1$. We obtain

$$C_\vartheta^{\mathrm{Cl}}(u_1, \ldots, u_d) := \left(\sum_{i=1}^d u_i^{-\vartheta} - (d-1) \right)^{-1/\vartheta},$$

cf. Example 4.30 for $d = 2$. They have the following properties:

(a) lower tail dependence, cf. Example 4.30,

(b) $\lim_{\vartheta \to 0} C_0^{\mathrm{Cl}}(u_1, \ldots, u_d) = \prod_{i=1}^d u_i$, which is the copula for independent random variables,

(c) $\lim_{\vartheta \to \infty} C_\vartheta^{\mathrm{Cl}}(u_1, \ldots, u_d) = \min\{u_1, \ldots, u_d\}$, which is the copula for comonotone random variables,

(d) symmetry in $u_1, \ldots, u_d$.

Kendall's τ can be computed from the generator of an Archimedean copula.

Theorem 4.43 *Let $X = (X_1, X_2)$ be a random vector with continuous marginals and Archimedean copula with generator φ. Then Kendall's τ equals*

$$\varrho_\tau(X_1, X_2) = 1 + 4 \int_0^1 \frac{\varphi(u)}{\varphi'(u)} du.$$

Example 4.44 1. For the Gumbel copula with parameter ϑ as in Examples 4.29 and 4.42, Kendall's τ equals $\varrho_\tau(\vartheta) = 1 - 1/\vartheta$ because

$$\begin{aligned} \varrho_\tau(\vartheta) &= 1 + 4 \int_0^1 \frac{(-\log u)(-u)}{\vartheta} du \\ &= 1 + \frac{4}{\vartheta} \left(\left[\frac{u^2}{2} \log u \right]_0^1 - \int_0^1 \frac{u^2}{2u} du \right) \\ &= 1 + \frac{4}{\vartheta} \left(0 - \frac{1}{4} \right). \end{aligned}$$

2. For the Clayton copula with parameter ϑ as in Examples 4.30 and 4.42, Kendall's τ equals $\varrho_\tau(\vartheta) = \vartheta/(\vartheta+2)$ because

$$\begin{aligned}\varrho_\tau(\vartheta) &= 1 + 4\int_0^1 \frac{u^{-\vartheta}-1}{-\vartheta u^{-\vartheta-1}}du \\ &= 1 + \frac{4}{\vartheta}\left[\frac{1}{\vartheta+2}u^{\vartheta+2} - \frac{u^2}{2}\right]_0^1 \\ &= 1 + \frac{4}{\vartheta}\left(\frac{1}{\vartheta+2} - \frac{1}{2}\right) \\ &= \frac{\vartheta}{\vartheta+2}.\end{aligned}$$

With regards to simulation, we consider Clayton copulas in the following example. Simulating Gumbel copulas is more difficult except for the case $d = 2$.

Example 4.45 (Simulation of Clayton copulas $C_\vartheta^{\mathrm{Cl}}$)

1. Simulate a realisation x of a Γ-distributed random variable with parameters $1/\vartheta$ and 1.
2. Simulate independent realisations $v_1, \ldots, v_d$ of a uniform distribution on $[0,1]$.
3. Set $u_k := (\frac{-\log(v_k)}{x} + 1)^{-1/\vartheta}$, $k = 1, \ldots, d$.
4. Now $u = (u_1, \ldots, u_d)$ is a realisation of a random vector with cdf $C_\vartheta^{\mathrm{Cl}}$.

4.3.5 Parameter estimation

The first step in applying copulas to risk management would be to choose an appropriate family of copulas e.g. following a graphical exploratory data analysis. Relevant criteria may be:

- Does the data seem to exhibit tail dependence?
- Is a possible asymmetry of X and $-X$ relevant for the application?
- Is it relevant to allow for different pairwise dependence of $X_1, \ldots, X_d$ in the case $d > 2$?

The next step is to estimate parameters in the chosen paramatric model, e.g. as suggested below. To this end, we assume that we have observed independent realisations of a random vector $X = (X_1, \ldots, X_d)$ with cdf F, marginal cdf's $F_1, \ldots, F_d$, and copula C. The goal is to estimate the parameters of C.

Step 1: Estimate Kendall's τ by $(\hat{\varrho}_\tau)_{ij} = \hat{\varrho}_\tau(X_i, X_j)$ for $i, j = 1, \ldots, d$, using the estimator from (4.1).

Step 2: From Step 1 we obtain an estimated copula having a similar tail behaviour as the data:

- In the case of a Gaussian copula C_R^{Ga} estimate R according to $\hat{R} = \sin(\pi(\hat{\varrho}_\tau)_{ij}/2)$, $i, j = 1, \ldots, d$, cf. (4.5). If R fails to be positive definite, replace it by a similar positive definite matrix.

- In the case of a t-copula $C_{\nu,R}^t$, R can be estimated as for the Gaussian copula. In order to estimate ν, replace the data $X_{(k)} = (X_{1,k}, \ldots, X_{d,k})$, $k = 1, \ldots, N$ by $U_{(k)} = (\hat{F}_1(X_{1,k}), \ldots, \hat{F}_d(X_{d,k}))$, $k = 1, \ldots, N$, where $\hat{F}_i$ denotes the empirical distribution function of $X_{i,k}$, $k = 1, \ldots, N$. These transformed random vectors $U_{(k)}, k = 1, \ldots, N$, are approximately i.i.d. with distribution function $C_{\nu,R}^t$. Since R has already been estimated, one may now try to estimate the remaining parameter ν by numerical maximisation of the log-likelihood.

- In case of a Gumbel copula $C_\vartheta^{\mathrm{Gu}}$, estimate ϑ by $\hat{\vartheta} := 1/(1 - (\hat{\varrho}_\tau)_{ij})$ with $i \neq j$, cf. 4.44(1). Note that $1/(1 - (\varrho_\tau)_{ij})$ does not depend on the chosen pair i, j in the presence of a Gumbel copula, so that a strong dependence of $\hat{\vartheta}$ on i, j may indicate inconsistency with the data.

- In case of a Clayton copula $C_\vartheta^{\mathrm{Cl}}$, one can proceed as with the Clayton copula, but with $\hat{\vartheta} := 2(\hat{\varrho}_\tau)_{ij}/(1 - (\hat{\varrho}_\tau)_{ij})$, cf. 4.44(2).

4.3.6 Comparison of multivariate models

A short summary of the multivariate models in these notes is provided in the table below.

Model/method	historical simulation	variance-covariance	1. multivariate normal distribution	2. normal mixture distribution, e.g. t	3. elliptical distribution, e.g. t	Copula		
						4. elliptical , e.g. t	Archimedean	
							5. Clayton	6. Gumbel
parameter estimation	not needed	empirical mean and empirical (co)variances	empirical mean and empirical (co)variances	model dependent	model dependent	model dependent, for t: Section 4.3.5 plus estimation of marginals	Sect. 4.3.5 plus estimation of marginals	Section 4.3.5 plus estimation of marginals
simulation	not needed, but can be done using the empirical law	if needed: $X = \mu + AZ$ with $Z \sim N_d(0, 1_d)$ and Cholesky decomposition $\Sigma = AA^\top$ cf. Ex. 4.18	for $X \sim N_d(\mu, \Sigma)$: $X = \mu + AZ$ with $Z \sim N_d(0, 1_d)$ and Cholesky decomposition $\Sigma = AA^\top$ cf. Ex. 4.18	Def. 4.10: $X = \mu + WAZ$	Rem. 4.17(6)	Ex. 4.37 plus simulation of marginals	Ex. 4.45 plus simulation of marginals	not discussed in these notes
VaR, ES	with empirical law	Equations (2.2, 2.3)	a) Monte Carlo simulation or b) linearisation	a) Monte Carlo simulation or b) linearisation	a) Monte Carlo simulation or b) linearisation	Monte Carlo simulation	Monte Carlo simulation	Monte Carlo simulation
Advantages	very simple, no model error except for iid assumption	very simple, not much data needed	simple	allows tail dependence	allows tail dependence	allows tail dependence, flexible marginals	allows tail dependence, flexible marginals, $\pm$-asymmetry possible	allows tail dependence, flexible marginals, $\pm$-asymmetry possible
Problems	needs a lot of data, extrapolation impossible	potentially huge model risk	often inappropriate	$\pm$-symmetry	$\pm$-symmetry	$\pm$-symmetry	one parameter copula, symmetry	one parameter copula, symmetry
Remarks								

Table 4.1: Overview over some multivariate models in risk management

Appendix A

Construction of estimators

Most of the statistical estimators in this course — if not all — rely essentially on just a few basic principles, which are summarized below. However, we do not justify rigorously that they make sense. Nor do we provide an extensive list of approaches how to come up with reasonable estimators.

1. *Replacing the unknown true with the empirical distribution*

   - Goal: to estimate the law P^X or a generalized moment $E(f(X)) = \int f dP^X$ from iid data $X_1, \ldots, X_n$ with law P^X
   - Idea: use the empirical law $\widehat{P}_n := \frac{1}{N} \sum_{n=1}^{N} \varepsilon_{X_i}$ as an estimator of P^X resp. use

   $$\widehat{E(f(X))} := \int f d\widehat{P}_n = \frac{1}{N} \sum_{n=1}^{N} f(X_n)$$

   as an estimator of $E(f(X)) = \int f dP^X$.
   - Justification: The law of large numbers resp. the Glivenko-Cantelli theorem yields consistency, i.e convergence of the estimator to the true quantity for $n \to \infty$.

2. *Plug-in estimators*

   - Goal: to estimate $f(c_1, \ldots, c_k)$ where $c_1, \ldots, c_k$ depend on the unknown law P^X and consistent estimators $(\widehat{C}_{i,n})_{n \in \mathbb{N}}$ are known for $c_i, i = 1, \ldots, k$
   - Idea: use $f(\widehat{C}_{1,n}, \ldots, \widehat{C}_{k,n})$ as an estimator for $f(c_1, \ldots, c_k)$.
   - Justification: if f is continuous, consistency $\widehat{C}_{i,n} \to c_i$ implies that $f(\widehat{C}_{1,n}, \ldots, \widehat{C}_{k,n}) \to f(c_1, \ldots, c_k)$ for $n \to \infty$, i.e. consistency of the estimator.

3. *Maximum likelihood estimator (MLE)*

   - Goal: to estimate the parameter vector ϑ of P^X based on independent or even dependent data $X = (X_1, \ldots, X_n)$ with law P^X

- Idea: take the maximizer of $\vartheta \mapsto \log(p(\vartheta; X_1, \ldots, X_n))$, where $x \to p(\vartheta; x)$ denotes the pdf of P^X.
- Justification: MLE's are often consistent and converging at the optimal rate.

4. *Resampling / Monte Carlo*

   - Goal: to estimate/compute $E(f(X)) = \int f dP^X$ if an estimator $\widehat{P}^X$ for the unknown law P^X is available such that iid samples with law $\widehat{P}^X$ can be simulated easily.
   - Idea: use
   $$\widehat{E(f(X))} := \frac{1}{N} \sum_{n=1}^{N} f(X_n)$$
   as an estimator of $E(f(X)) = \int f dP^X$, where $X_1, \ldots, X_N$ denotes a simulated iid sample from law $\widehat{P}^X$.
   - If $\widehat{P}^X = P^X$ (at least approximately), then $\widehat{E(f(X))} \approx E(f(X))$ follows for sufficiently large N from the law of large numbers.

Bibliography

[HL07] H Hult and F Lindskog. *Mathematical modeling and statistical methods for risk management*. Lecture Notes, Stockholm, 2007.

[MFE10] A. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press, 2010.