Quantitative Risk Management

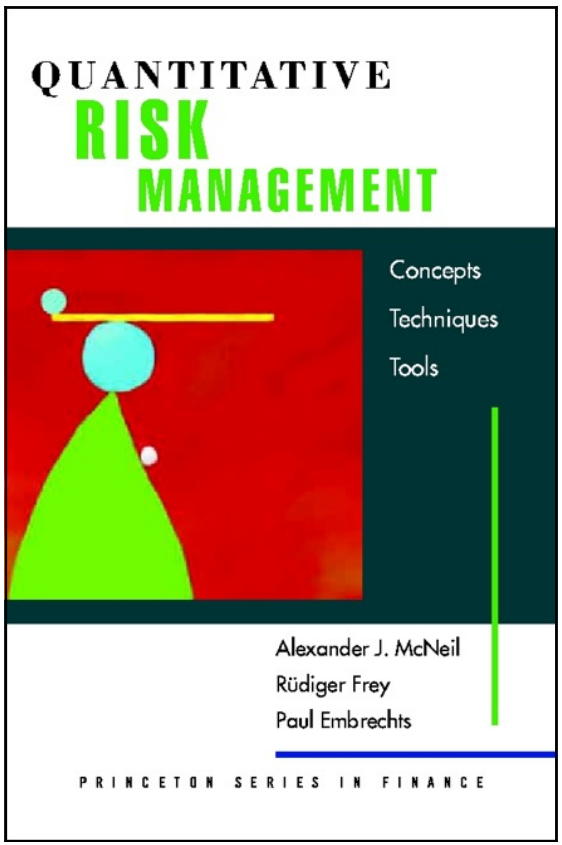
https://www.qrmtutorial.org

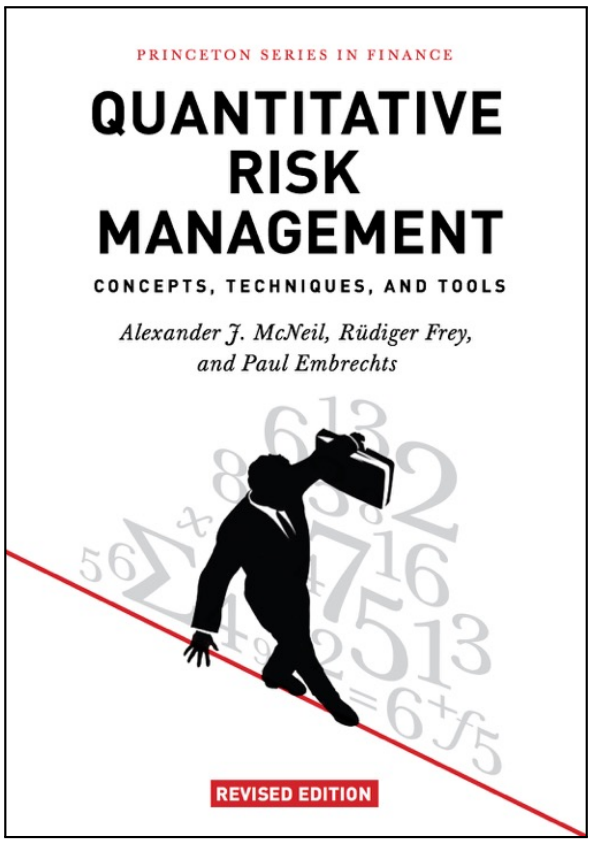
Last update: 2020-01-03

P. Embrechts, R. Frey, M. Hofert, A. J. McNeil

Course information

- Website: `https://www.qrmtutorial.org`
- Book: A. J. McNeil, R. Frey, P. Embrechts
  Quantitative Risk Management (1st edition: 2005; revised edition: 2015)





Overview

1 Risk in perspective

1.1 Risk

1.2 Why manage financial risk?

1.3 Quantitative Risk Management

1.1 Risk

- The Concise Oxford English Dictionary: "hazard, a chance of bad consequences, loss or exposure to mischance".
- McNeil, Frey, and Embrechts (2005): "any event or action that may adversely affect an organization's ability to achieve its objectives and execute its strategies".
- No single one-sentence definition captures all aspects of risk.
  For us: *risk* = chance of loss $\Rightarrow$ randomness

1.1.1 Risk and randomness

- We will mostly model situations in which an investor holds today an asset with an uncertain future value.
- We use probabilistic notions (random variables, random vectors, distributions, stochastic processes) and statistical tools. In particular, we assume to work on a *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$; see Kolmogorov (1933).

1.1.2 Financial Risk

There are various types of risks. We focus on (those affected by regulation):

Market risk Risk of loss in a financial position due to changes in the underlying components (e.g. stock/bond/commodity prices)

Credit risk Risk of a counterparty failing to meet its obligations (default), i.e. the risk of not receiving promised repayments (e.g. loans/bonds).

Operational risk (OpRisk) Risk of loss resulting from inadequate or failed internal processes, people and systems or from external events (e.g. fraud, fat-finger trades, earthquakes).

There are many other types of risks such as liquidity risk, underwriting risk, or model risk (the risk of using a misspecified or inappropriate model for measuring risk; model risk is always present to some degree).

1.1.3 Measurement and management

Risk measurement

- Suppose we hold a portfolio of d investments with weights $w_1, \ldots, w_d$. Let X_j denote the change in value of the jth investment. The change in value – *profit and loss (P&L)* – of the portfolio over a given holding period is then

$$X = \sum_{j=1}^{d} w_j X_j.$$

  Measuring the risk now consists of determining the *distribution function* F (or functionals of it, e.g. mean, variance, α-quantiles $F^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$).

- To this end, we need a properly calibrated joint model for $\boldsymbol{X} = (X_1, \ldots, X_d)$. Statistical estimates of F or one of its functionals are obtained based on historical observations of this model.

Risk management

- What is RM? Kloman (1990) writes:

  "RM is a discipline for living with the possibility that future events may cause adverse effects."

  $\Rightarrow$ It is about ensuring resilience to future events.

- Note that financial firms are not passive/defensive towards risk, banks and insurers actively/willingly take risks because they seek a return. RM thus belongs to their core competence.

- What does managing risks involve?
  - Determine the capital to hold to absorb losses, both for *regulatory purposes* (to comply with regulators) and *economic capital* purposes (to survive as a company).
  - Ensuring portfolios are well diversified.
  - Optimizing portfolios according to risk-return considerations.

1.2 Why manage financial risk?

- Society (single customers and as a whole (systemic risk)) relies on the stability of the banking and insurance system.
- This is related to *systemic importance* of the company in question (size and connectivity to other firms). Considering some firms as too big to fail creates a moral hazard (should be avoided!) since the management of such a firm may take more risk knowing that it would be bailed out in a crisis.
- Better risk management can reduce the risk of company failure and protect customers and policyholders. However, regulation must be designed with care and should not promote herding, procyclical behaviour or other forms of endogenous risk that could result in a systemic crisis.

1.3 Quantitative Risk Management

1.3.1 The Q in QRM

- We treat QRM as a quantitative science using the language of mathematics in general, and probability and statistics in particular.
- Probability and statistics provide us with a suitable language and with appropriate concepts for describing financial risks.
- We also point out assumptions and limitations of the methodology used.
- The Q in QRM is an essential part of the RM process. We believe it remains (if applied correctly and honestly) a part of the solution to managing risk (not the problem). See also Shreve (2008):

  "Don't blame the quants. Hire good ones instead and listen to them."

1.3.2 The nature of the challenge

- Our approach to QRM has two main strands:
  - Put current practice onto a firmer mathematical ground;
  - Put together techniques and tools which go beyond current practice and address some of the deficiencies.
- In particular, some of the challenges of QRM are:
  - Extremes matter.
  - Interdependence and concentration of risks.
  - The problem of scale (models for all risk factors may not be feasible).
  - Interdisciplinarity.
  - Communication and education.

2 Basic concepts in risk management

2.1 Risk management for a financial firm

2.2 Modelling value and value change

2.3 Risk measurement

2.1 Risk management for a financial firm

2.1.1 Assets, liabilities and the balance sheet

The risks of a firm (here: bank) can be understood from its balance sheet:

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

A stylized balance sheet for an insurer is:

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

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

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

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

2.1.2 Risks faced by a financial firm

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

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

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

2.1.3 Capital

- There are different notions of capital. One distinguishes:

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

Regulatory capital
- Capital required according to regulatory rules;

- For European insurance firms: Minimum (MCR) and solvency capital requirements (SCR);
- A regulatory framework also specifies the capital quality. One distinguishes *Tier 1 capital* (i.e. shareholder capital + retained earnings; can act in full as buffer) and *Tier 2 capital* (includes other positions on the balance sheet).

Economic capital

- Capital required to control the probability of becoming insolvent (typically over one year);
- Internal assessment of risk capital;
- Aims at a holistic view (assets and liabilities) and works with fair values of balance sheet items.

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

2.2 Modelling value and value change

2.2.1 Mapping of risks

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

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

  1) the portfolio composition remains fixed over Δt;

  2) there are no intermediate payments during Δt

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

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

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

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

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

(as QRM is mainly concerned with losses).

Remark 2.1

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

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

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

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

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

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

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

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

We see that the loss df is determined by the loss df of $\boldsymbol{X}_{t+1}$. We will thus also write $L_{t+1} = L(\boldsymbol{X}_{t+1})$, where $L(\boldsymbol{x}) = -(f(t+1, \boldsymbol{Z}_t + \boldsymbol{x}) - f(t, \boldsymbol{Z}_t))$ is known as *loss operator*.

- If f is differentiable, its first-order (Taylor) approximation ($f(\boldsymbol{y}) \approx f(\boldsymbol{y}_0) + \nabla f(\boldsymbol{y}_0)'(\boldsymbol{y} - \boldsymbol{y}_0)$ for $\boldsymbol{y} = (t+1, Z_{t,1} + X_{t+1,1}, \ldots, Z_{t,d} + X_{t+1,d}$ and $\boldsymbol{y}_0 = (t, Z_{t,1}, \ldots, Z_{t,d})$) is

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

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

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

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

Example 2.2 (Stock portfolio)

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

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

- The one-period ahead loss is then given by

$$L_{t+1} = -(V_{t+1} - V_t) = -\sum_{j=1}^{d} \lambda_j (e^{Z_{t,j}+X_{t+1,j}} - e^{Z_{t,j}})$$

$$= -\sum_{j=1}^{d} \lambda_j e^{Z_{t,j}} (e^{X_{t+1,j}} - 1) = -\sum_{j=1}^{d} \underbrace{\lambda_j S_{t,j}}_{=: \tilde{w}_{t,j}} (e^{X_{t+1,j}} - 1) \quad (1)$$

  which is non-linear in $X_{t+1,j}$ (here: $L(\boldsymbol{x}) = -\sum_{j=1}^{d} \tilde{w}_{t,j}(e^{x_j} - 1)$).

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

$$L_{t+1}^{\Delta} = -\left(f_t(t, \boldsymbol{Z}_t) + \sum_{j=1}^{d} f_{z_j}(t, \boldsymbol{Z}_t) X_{t+1,j} \right) = -\left(0 + \sum_{j=1}^{d} \tilde{w}_{t,j} X_{t+1,j} \right)$$
$$= -\tilde{\boldsymbol{w}}_t' \boldsymbol{X}_{t+1}.$$

- Note that $L_{t+1}^{\Delta} = -(c_t + \boldsymbol{b}_t' \boldsymbol{X}_{t+1})$ for $c_t = 0$ and $\boldsymbol{b}_t = \tilde{\boldsymbol{w}}_t$.
- If $\boldsymbol{\mu} = \mathbb{E}\boldsymbol{X}_{t+1}$ and $\Sigma = \operatorname{cov} \boldsymbol{X}_{t+1}$ are known, then expectation and variance of the (linearized) one-period ahead loss are

$$\mathbb{E} L_{t+1}^{\Delta} = -\sum_{j=1}^{d} \tilde{w}_{t,j} \mathbb{E}(X_{t+1,j}) = -\tilde{\boldsymbol{w}}_t' \boldsymbol{\mu},$$
$$\operatorname{var} L_{t+1}^{\Delta} = \operatorname{var}(\tilde{\boldsymbol{w}}_t' \boldsymbol{X}_{t+1}) = \tilde{\boldsymbol{w}}_t' \operatorname{cov}(\boldsymbol{X}_{t+1}) \tilde{\boldsymbol{w}}_t = \tilde{\boldsymbol{w}}_t' \Sigma \tilde{\boldsymbol{w}}_t.$$

- If $\boldsymbol{X}_{t+1}$ is multivariate normal, then $L_{t+1}^{\Delta} \sim \mathrm{N}(-\tilde{\boldsymbol{w}}_t' \boldsymbol{\mu}, \tilde{\boldsymbol{w}}_t' \Sigma \tilde{\boldsymbol{w}}_t)$.

Example 2.3 (European call option)

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

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

where

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

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

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

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

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

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

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

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

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

2.2.2 Loss distributions

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

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

There are three general methods to approach these challenges.

1) Analytical method

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

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

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

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

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

Advantages:
- $F_{L^\Delta_{t+1}}$ explicit ($\Rightarrow$ typically explicit risk measures)
- Easy to implement (unless d extremely large)

Drawbacks: Assumption 1 is unlikely to be realistic for daily (probably also weekly/monthly) data (see later).

2) Historical simulation

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

$$\hat{F}_{L_{t+1},n}(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{L_{t-i+1} \leq x\}}, \quad x \in \mathbb{R}, \tag{3}$$

based on

$$L_k = L(\boldsymbol{X}_k) = -(f(t+1, \boldsymbol{Z}_t + \boldsymbol{X}_k) - f(t, \boldsymbol{Z}_t)), \tag{4}$$

$k \in \{t-n+1, \ldots, t\}$. $L_{t-n+1}, \ldots, L_t$ show what would happen to the current portfolio if the past n risk-factor changes were to recur.

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

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

3) Monte Carlo method

Idea: Take any model for $\boldsymbol{X}_{t+1}$, simulate $\boldsymbol{X}_{t+1}$, compute the corresponding losses as in (4) and estimate $F_{L_{t+1}}$ (typically via edf as in (3)).

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

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

So-called *economic scenario generators* used in insurance also fall under the heading of Monte Carlo methods.

2.3 Risk measurement

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

2.3.1 Approaches to risk measurement

Existing risk measurement approaches grouped into three categories:

1) **Notional-amount approach**
   - oldest approach; "standardized approaches" of Basel II (e.g. OpRisk)
   - *risk of a portfolio* = summed notional values of the securities times their riskiness factor.

2) **Risk measures based on loss distributions**
   - Most modern risk measures are characteristics of the underlying loss distribution over some predetermined time horizon Δt.
   - Examples: variance, value-at-risk, expected shortfall (see later)
   - Advantages:
     - Makes sense on all levels (from single portfolios to the overall position of a financial institution).
     - Loss distributions reflect netting and diversification.

3) Scenario-based risk measures

- Typically considered in stress testing.
- One considers possible future risk-factor changes (*scenarios*; e.g. a 20% drop in a market index).
- *Risk of a portfolio* = maximum (weighted) loss under all scenarios.
- If $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ denote the risk-factor changes (scenarios) with corresponding weights $\boldsymbol{w} = (w_1, \ldots, w_n)$, the risk is $\psi_{\mathcal{X},\boldsymbol{w}} = \max_{1 \le i \le n}\{w_i L(\boldsymbol{x}_i)\}$, where $L(\boldsymbol{x})$ denotes the loss the portfolio would suffer if the hypothetical scenario $\boldsymbol{x}$ were to occur. Many risk measures are of this form; see *CME SPAN: Standard Portfolio Analysis of Risk* (2010).
- Mathematical interpretation:
  - Assume $L(\mathbf{0}) = 0$ (okay if Δt small) and $w_i \in [0, 1]\ \forall\, i$.

- $w_i L(\boldsymbol{x}_i) = w_i L(\boldsymbol{x}_i) + (1 - w_i) L(\mathbf{0}) = \mathbb{E}_{\mathbb{P}_i}(L(\boldsymbol{X}))$ where $\boldsymbol{X} \sim \mathbb{P}_i$ and $\mathbb{P}_i$ is such that $\mathbb{P}_i(\boldsymbol{X} = \boldsymbol{x}_i) = w_i$ and $\mathbb{P}_i(\boldsymbol{X} = \mathbf{0}) = 1 - w_i$.

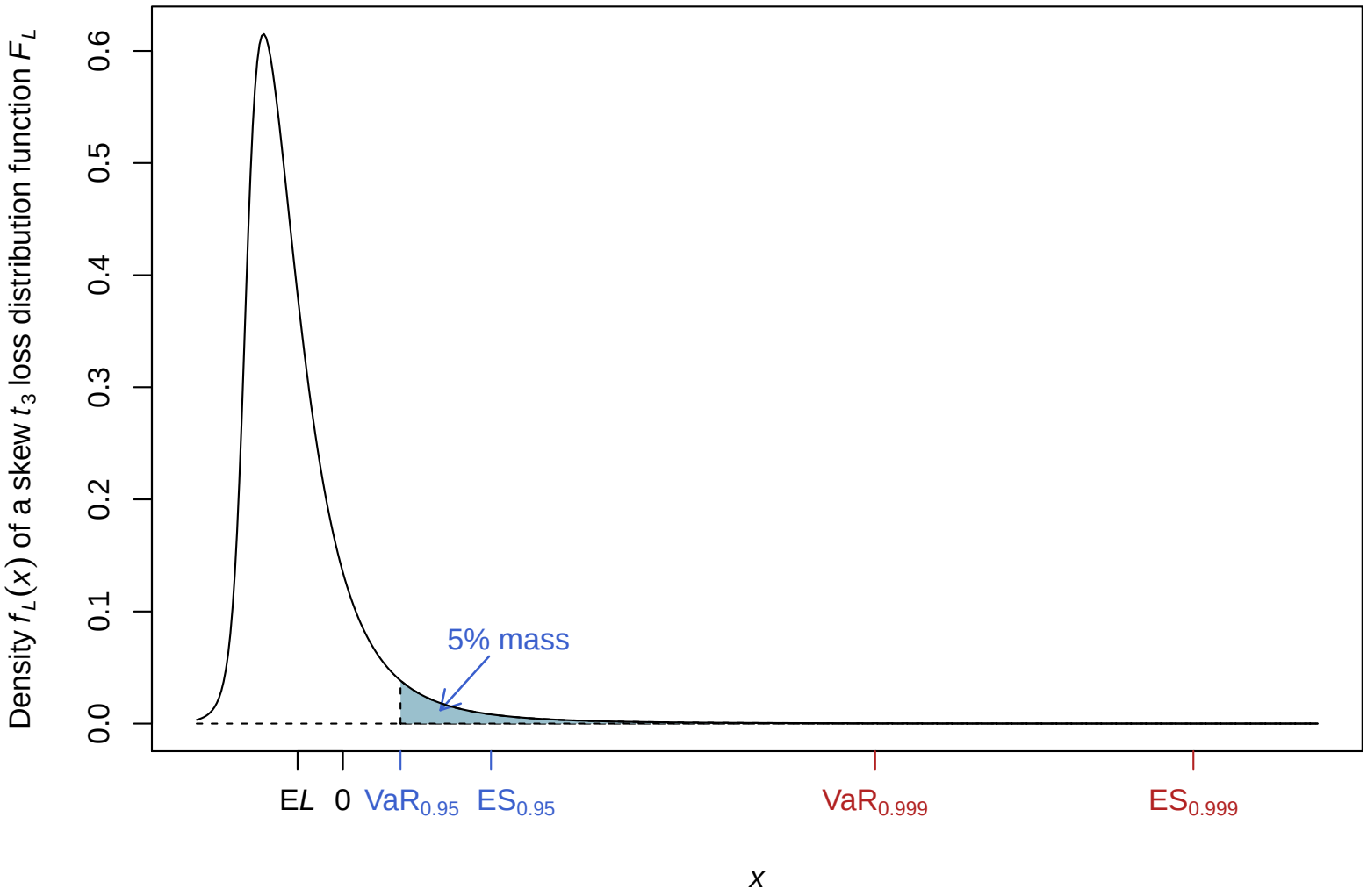
Therefore, $\psi_{\mathcal{X},\boldsymbol{w}} = \max\{\mathbb{E}_{\mathbb{P}}(L(\boldsymbol{X})) : \boldsymbol{X} \sim \mathbb{P} \in \{\mathbb{P}_1, \dots, \mathbb{P}_n\}\}$. Such a risk measure is known as a *generalized scenario*.

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

2.3.2 Value-at-risk

Definition 2.4 (Value-at-risk)
For a loss $L \sim F_L$, *value-at-risk* (VaR) at confidence level $\alpha \in (0, 1)$ is defined by $\mathrm{VaR}_\alpha = \mathrm{VaR}_\alpha(L) = F_L^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F_L(x) \geq \alpha\}$.

- VaR_α is simply the α-quantile of F_L. As such, $F_L(\mathrm{VaR}_\alpha(L)) = F_L(F_L^{\leftarrow}(\alpha)) \geq \alpha$ and $F_L(x) < \alpha$ for all $x < \mathrm{VaR}_\alpha(L)$.
- $\mathrm{VaR}_\alpha(L) = \inf\{x \in \mathbb{R} : \bar{F}_L(x) \leq 1 - \alpha\}$ $(\bar{F}_L(x) = 1 - F_L(x))$, so VaR_α is the smallest loss which is exceeded with prob. at most $1 - \alpha$.
- Known since 1994: Weatherstone 4^{15} report (J.P. Morgan; RiskMetrics)
- VaR is the most widely used risk measure (by Basel II or Solvency II)
- $\mathrm{VaR}_\alpha(L)$ is not a what if risk measure: It does not provide information about the severity of losses which occur with probability $\leq 1 - \alpha$.


Density $f_L(x)$ of a skew t_3 loss distribution function F_L
0.0
0.1
0.2
0.3
0.4
0.5
0.6
5% mass
EL
0
$\text{VaR}_{0.95}$
$\text{ES}_{0.95}$
$\text{VaR}_{0.999}$
$\text{ES}_{0.999}$
x

Interlude: Generalized inverses

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

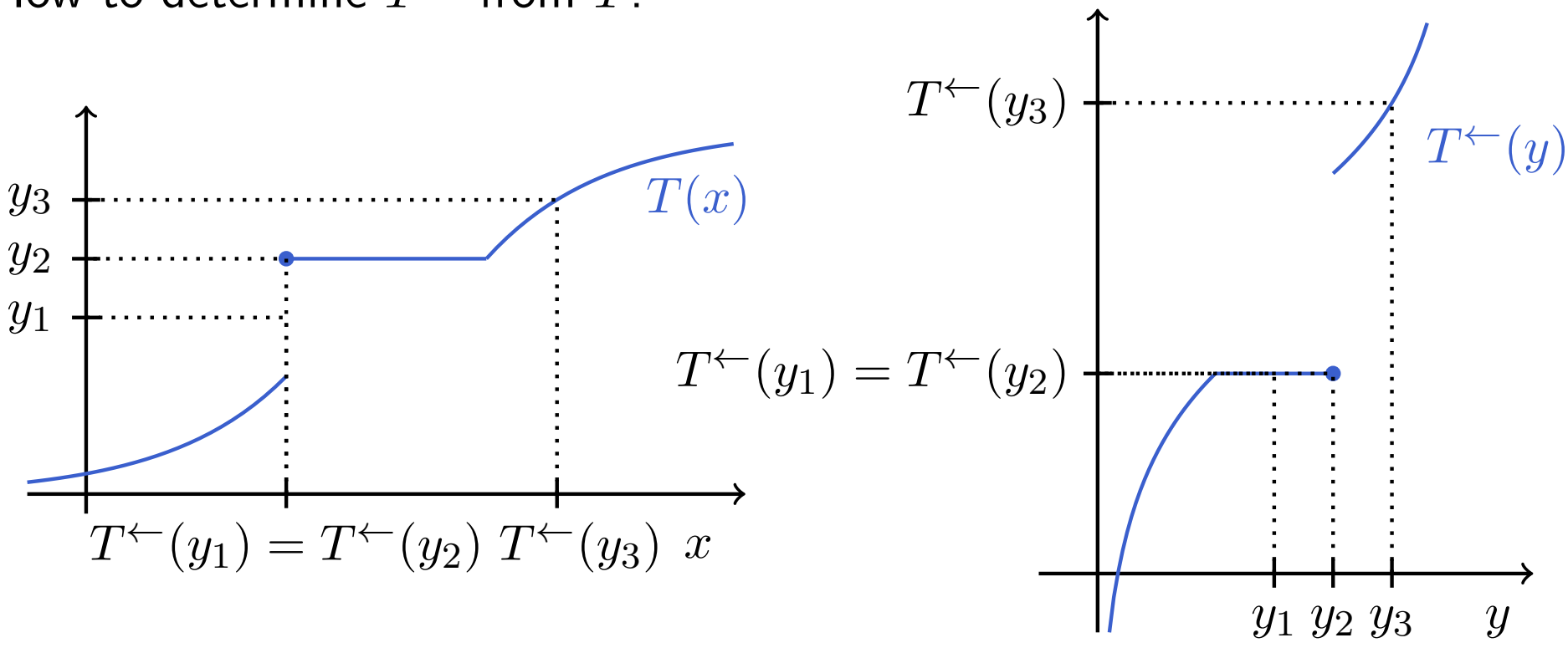
Definition 2.5 (Generalized inverse)
For any increasing function $T : \mathbb{R} \to \mathbb{R}$, with $T(-\infty) = \lim_{x \downarrow -\infty} T(x)$ and $T(\infty) = \lim_{x \uparrow \infty} T(x)$, the *generalized inverse* $T^{\leftarrow} : \mathbb{R} \to \bar{\mathbb{R}} = [-\infty, \infty]$ of T is defined by

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

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

- If T is continuous and $\uparrow$, then $T^{\leftarrow} \equiv T^{-1}$ (ordinary inverse).
- There are rules for working with $T^{\leftarrow}$ (often, not always) similar to T^{-1}.

How to determine $T^{\leftarrow}$ from T:



- Flat parts (jumps) of T correspond to jumps (flat parts) of $T^{\leftarrow}$.
- Assume T to be a df and $L \sim T$.
  - What is the probability that L falls in the region where T is flat?
  - What is $\mathbb{P}(L = T^{\leftarrow}(y_1))$?

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

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

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

This implies that

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

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

2) Let $L \sim t_\nu(\mu,\sigma^2)$, so $(L-\mu)/\sigma \sim t_\nu = t_\nu(0,1)$ and thus, as above,

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

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

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

Furthermore, if $\nu > 1$, $\mathbb{E}X$ exists and $\mathbb{E}X = 0$. If $\nu > 2$, then $\mathrm{var}\, X$ exists and $\mathrm{var}\, X = \frac{\nu}{\nu-2}$; in particular, $Z = \sqrt{\frac{\nu-2}{\nu}}X \sim t_\nu(0, \frac{\nu-2}{\nu})$ has $\mathrm{var}\, Z = 1$.

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

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

2.3.3 VaR in risk capital calculations

VaR in regulatory capital calculations for the trading book

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

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

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

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

2.3.4 Expected shortfall

Definition 2.7 (Expected shortfall)
For a loss $L \sim F_L$ with $\mathbb{E}(L_+) = \mathbb{E}(\max\{L, 0\}) < \infty$, *expected shortfall* ($\mathrm{ES}$) at confidence level $\alpha \in (0, 1)$ is defined by

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

- ES_α is the average of VaR_u over all $u \geq \alpha \Rightarrow \mathrm{ES}_\alpha \geq \mathrm{VaR}_\alpha$.
- ES_α looks further into the tail of F_L, it is a "what if" risk measure (VaR_α is frequency-based; ES_α is severity-based). This also becomes clear from the following result which shows that under continuity, expected shortfall equals *conditional tail expectation* or *tail value-at-risk*.

Proposition 2.8 ($\mathrm{ES}_\alpha(L)$ under continuity)
If F_L is continuous, $\mathrm{ES}_\alpha(L) = \mathbb{E}(L \,|\, L > \mathrm{VaR}_\alpha(L))$.

Proof. If F_L is continuous, $F_L(\mathrm{VaR}_\alpha(L)) = F_L(F_L^{\leftarrow}(\alpha)) = \alpha$ and thus, for all $x \geq \mathrm{VaR}_\alpha(L)$,

$$\begin{aligned} F_{L|L>\mathrm{VaR}_\alpha(L)}(x) &= \mathbb{P}(L \leq x \,|\, L > \mathrm{VaR}_\alpha(L)) \\ &= \frac{\mathbb{P}(L \leq x,\ L > \mathrm{VaR}_\alpha(L))}{\mathbb{P}(L > \mathrm{VaR}_\alpha(L))} = \frac{\mathbb{P}(\mathrm{VaR}_\alpha(L) < L \leq x)}{\mathbb{P}(L > \mathrm{VaR}_\alpha(L))} \\ &= \frac{F_L(x) - F_L(\mathrm{VaR}_\alpha(L))}{1 - F_L(\mathrm{VaR}_\alpha(L))} = \frac{F_L(x) - \alpha}{1 - \alpha}. \end{aligned}$$

Since $\mathrm{d}F_{L|L>\mathrm{VaR}_\alpha(L)}(x) = \mathrm{d}F_L(x)/(1-\alpha)$,

$$\begin{aligned} \mathbb{E}(L \,|\, L > \mathrm{VaR}_\alpha(L)) &= \int_{\mathrm{VaR}_\alpha(L)}^{\infty} x \,\mathrm{d}F_{L|L>\mathrm{VaR}_\alpha(L)}(x) \\ &= \frac{1}{1-\alpha} \int_{\mathrm{VaR}_\alpha(L)}^{\infty} x \,\mathrm{d}F_L(x) = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_z(L) \,\mathrm{d}z \\ &= \mathrm{ES}_\alpha(L), \end{aligned}$$

where we substituted $x = \mathrm{VaR}_z(L) = F_L^{\leftarrow}(z)$ (so $F_L(x) = z$, $\mathrm{d}F_L(x) = \mathrm{d}z$). □

- ES_α is more difficult to estimate and backtest than VaR_α (the variance of estimators is typically larger; larger sample size required).
- $\mathrm{ES}_\alpha(L) < \infty$ requires $\mathbb{E}(L_+) < \infty$.
- Subadditivity and elicitability. One can show:
  - In contrast to VaR_α, ES_α is subadditive (more later).
  - In contrast to ES_α, VaR_α exists if $\mathbb{E}|L| = \infty$ and is elicitable (i.e. minimizes some expected functional (scoring function); see Gneiting (2011). This can be used for backtesting, comparing risk measures).

Example 2.9 (A comparison of VaR and ES for stock returns)

- Consider Example 2.2 with a 1-stock portfolio and $V_t = S_t = 10\,000$. In this case, $L^{\Delta}_{t+1} = -S_t X_{t+1}$, where $X_{t+1} = \log(S_{t+1}/S_t)$.
- Let $\sigma = 0.2/\sqrt{250}$ (annualized volatility of 20%) and assume

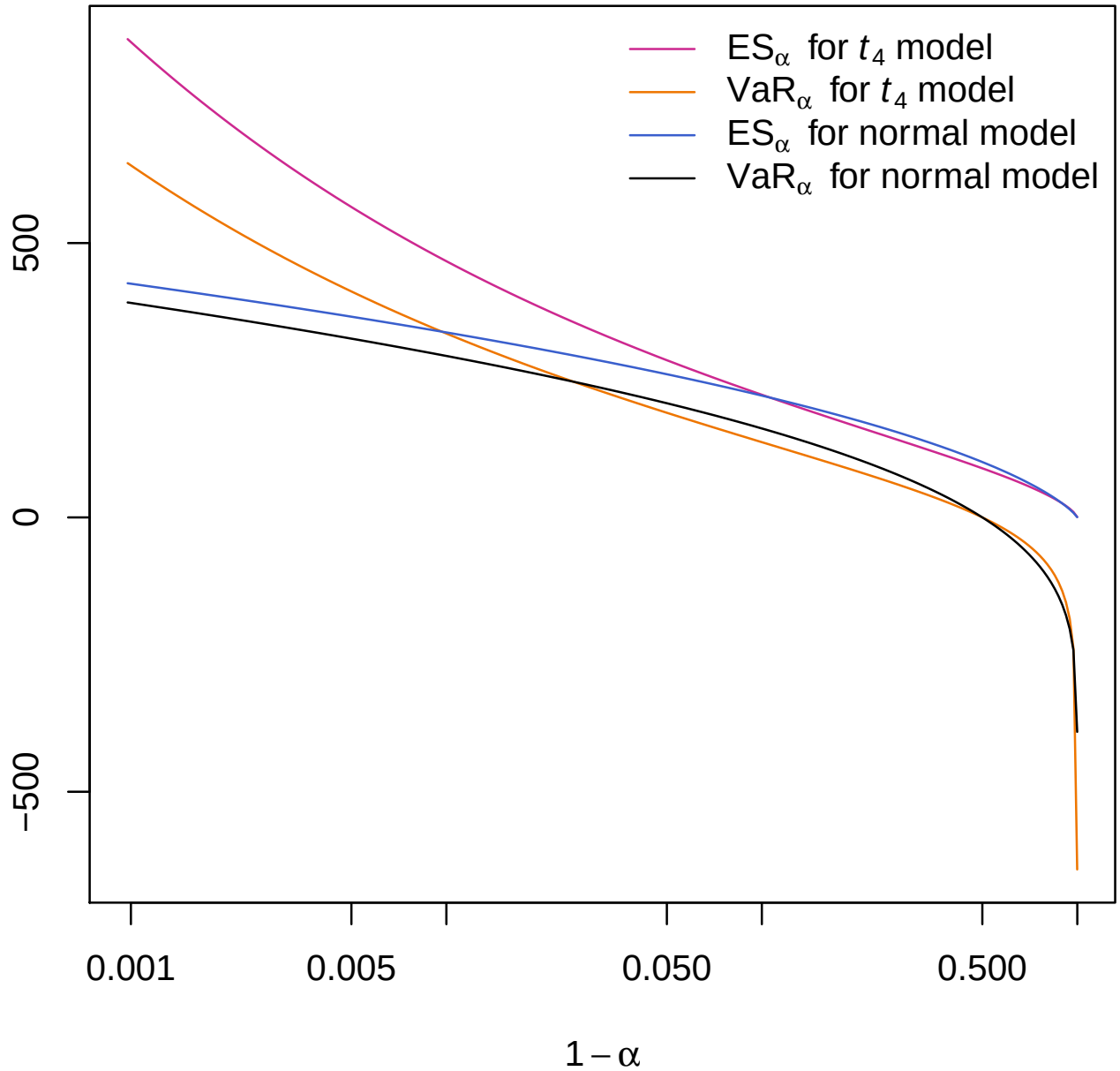
  1) $X_{t+1} \sim \mathrm{N}(0, \sigma^2) \Rightarrow L^{\Delta}_{t+1} \sim \mathrm{N}(0, S_t^2\sigma^2)$;

  2) $X_{t+1} \sim t_\nu(0, \sigma^2\frac{\nu-2}{\nu})$, $\nu > 2$ (so that $\mathrm{var}\, X_{t+1} = \sigma^2$, too). Then

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

  $\Rightarrow L^{\Delta}_{t+1} = -S_t\sqrt{\sigma^2\frac{\nu-2}{\nu}}Y \sim t_\nu(0, S_t^2\sigma^2\frac{\nu-2}{\nu})$ (so $\mathrm{var}(L^{\Delta}_{t+1}) = S_t^2\sigma^2$, too).

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



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

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

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

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

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

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

By linearity (or see soon), $L\sim\mathrm{N}(\mu,\sigma^2)$ has expected shortfall

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

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

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

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

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

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

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

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

2.3.5 Coherent and convex risk measures

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

  1) **Elements of $\mathcal{M}$ are net asset values** V_{t+1}: $\tilde{\varrho}(V_{t+1})$ denotes the capital to be added to a position with future value V_{t+1} to make it acceptable to a regulator.

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

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

Axioms of coherence

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

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

Criticism: None

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

Interpr.:

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

Criticism: Most people believe this to be reasonable.

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

Interpr.:

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

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

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

Interpr.: (or motivation): For $L_1 = \cdots = L_n = L$, subadditivity implies $\varrho(nL) \leq n\varrho(L)$, but there is no diversification, so equality should hold.

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

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

Coherent risk measures are convex. The converse is not true in general (but for positive homogeneous risk measures ϱ).

Example 2.12 (Coherence of generalized scenario risk measures)
Let $L(\boldsymbol{x})$ denote the hypothetical loss under scenario $\boldsymbol{x}$ (risk-factor change). The generalized scenario risk measure

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

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

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

Remark 2.13
One can show that all coherent risk measures can be represented as generalized scenarios via $\varrho(L) = \sup\{\mathbb{E}_{\mathbb{P}}(L) : \mathbb{P} \in \mathcal{P}\}$ for a suitable set $\mathcal{P}$ of probability measures.

Theorem 2.14 (Coherence of ES)

ES is a coherent risk measure.

Proof. Monotonicity, translation invariance and positive homogeneity follow from VaR. Subadditivity is more involved but can be shown. □

Superadditivity scenarios for VaR

Under the following scenarios, VaR_α is typically superadditive:

1) L_1, L_2 have skewed distributions;

2) Independent, light-tailed L_1, L_2 and small α;

3) L_1, L_2 have special dependence;

4) L_1, L_2 have heavy tailed distributions.

Let's have a look at examples for 1), 2) and 4); for 3), see later.

Example 2.15 (Skewed loss distributions)

Consider two independent losses of the form

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

Set $\alpha = 0.99$. Then $\mathrm{VaR}_\alpha(L_j) = -5$, $j \in \{1,2\}$. The loss $L_1 + L_2$ is

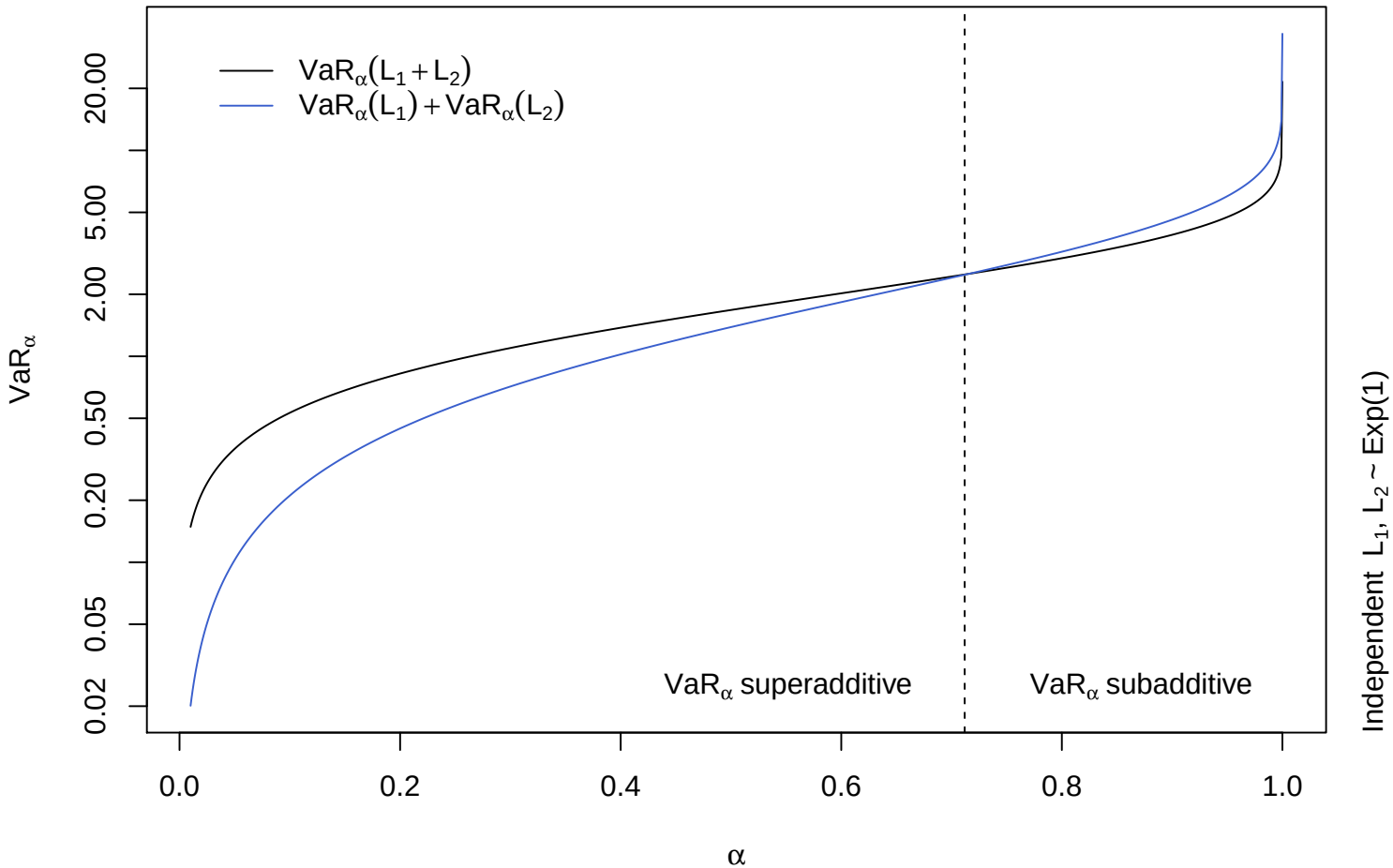
$$L_1 + L_2 = \begin{cases} -10, & \text{with prob. } (1-p)^2 = 0.982081, \\ 95, & \text{with prob. } 2p(1-p) = 0.017838, \\ 200, & \text{with prob. } p^2 = 0.000081. \end{cases}$$

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

- For d such losses, one can show that VaR_α is superadditive if and only if $(1-p)^d < \alpha \leq 1-p$.
- From the money lender's (investor) view, the losses could be two independently defaultable zero-coupon bonds (maturity $T = 1$ y, face value 100, interest 5%, default probability $p = 0.009$, no recovery).

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

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



One can show that (independently of λ) VaR_α is superadditive if and only if $(1-\alpha)(1-2\log(1-\alpha)) > 1$.

Example 2.17 (Heavy tailed loss distributions)

Let $L_1, L_2 \overset{\text{ind.}}{\sim} F(x) = 1 - x^{-1/2}$, $x \in [1, \infty)$. One can show (via density convolution formula; tedious!) that $F_{L_1+L_2}(x) = 1 - 2\sqrt{x-1}/x$, $x \geq 2$. By solving a quadratic equation one obtains that VaR_α is superadditive for all $\alpha \in (0, 1)$.

Remark 2.18 (Special case of comonotone risks)

- $L_1 \overset{\text{a.s.}}{=} L_2$ (special case of "comonotonicity") does not lead to the largest $\mathrm{VaR}_\alpha(L_1 + L_2)$ since $\mathrm{VaR}_\alpha(L_1 + L_2) = \mathrm{VaR}_\alpha(2L_1) = 2\,\mathrm{VaR}_\alpha(L_1) = \mathrm{VaR}_\alpha(L_1) + \mathrm{VaR}_\alpha(L_2)$, so "only" equality (whereas all above scenarios produced ">"). All previous examples thus gave a larger VaR under independence than comonotonicity!
- ES_α is subadditive (see Theorem 2.14) and comonotone additive (same idea as for VaR_α) and thus largest under comonotonicity.
- VaR_α is subadditive (so coherent) for all elliptical distributitions (strictly including the multivariate normal and t) when $\alpha \geq 1/2$; see later.

3 Empirical properties of financial data

3.1 Stylized facts of financial return series

3.2 Multivariate stylized facts

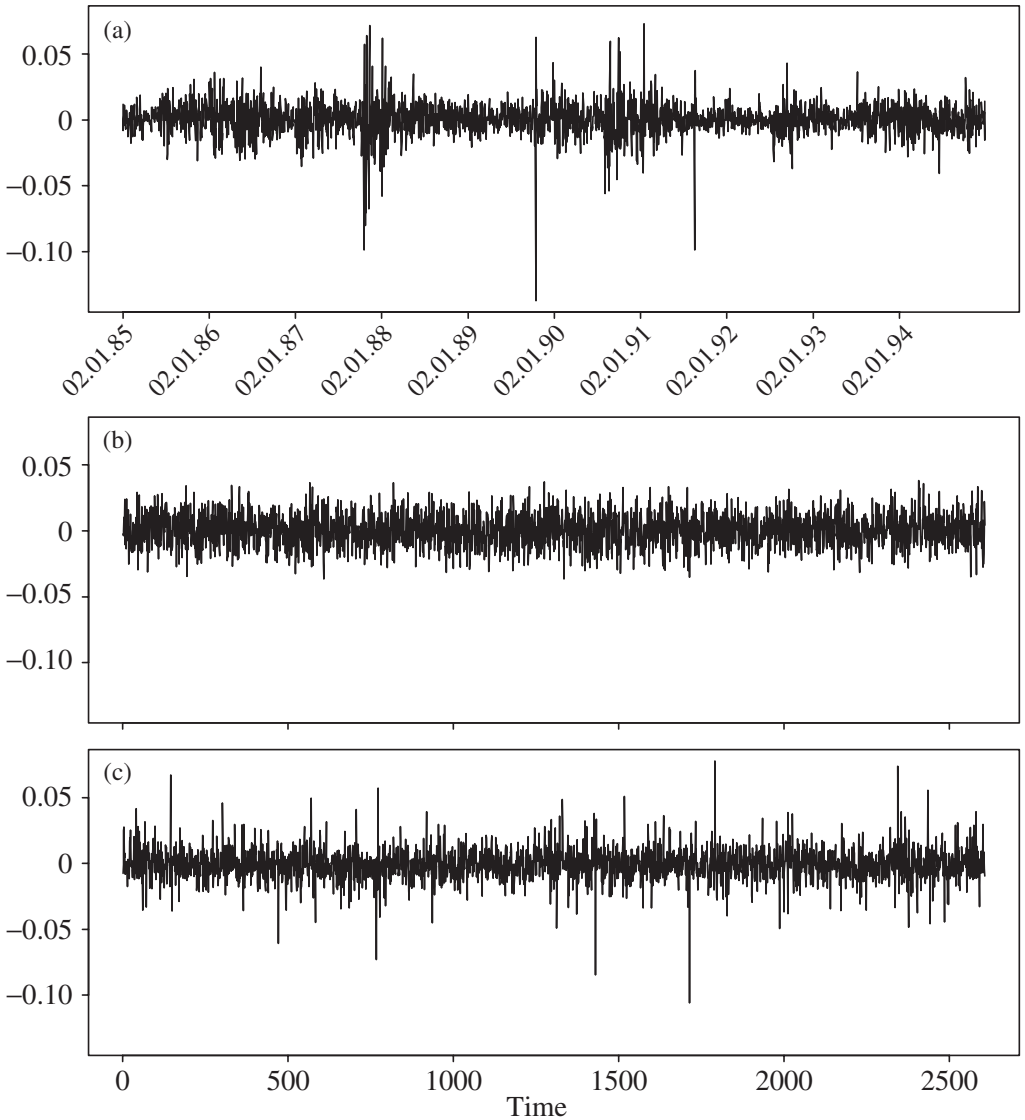
3.1 Stylized facts of financial return series

- Stylized facts are a collection of empirical observations and related inferences, which apply to many time series of risk-factor changes (e.g. log-returns on equities, indices, exchange rates, commodity prices).
- The best-known stylized facts apply to daily log-returns (also to intra-daily, weekly, monthly). Tick-by-tick (high-frequency) data have their own stylized facts (not discussed here) and annual return (low-frequency) data are more difficult to investigate (data sparseness; non-stationarity).
- Consider discrete-time risk-factor changes $X_t = Z_t - Z_{t-1}$ for a log-price or rate $Z_t = \log S_t$. In this case

$$X_t = \log(S_t/S_{t-1}) \approx S_t/S_{t-1} - 1 = (S_t - S_{t-1})/S_{t-1};$$

the former is often called a *(log-)return*, the latter a *simple return*.

3.1.1 Volatility Clustering

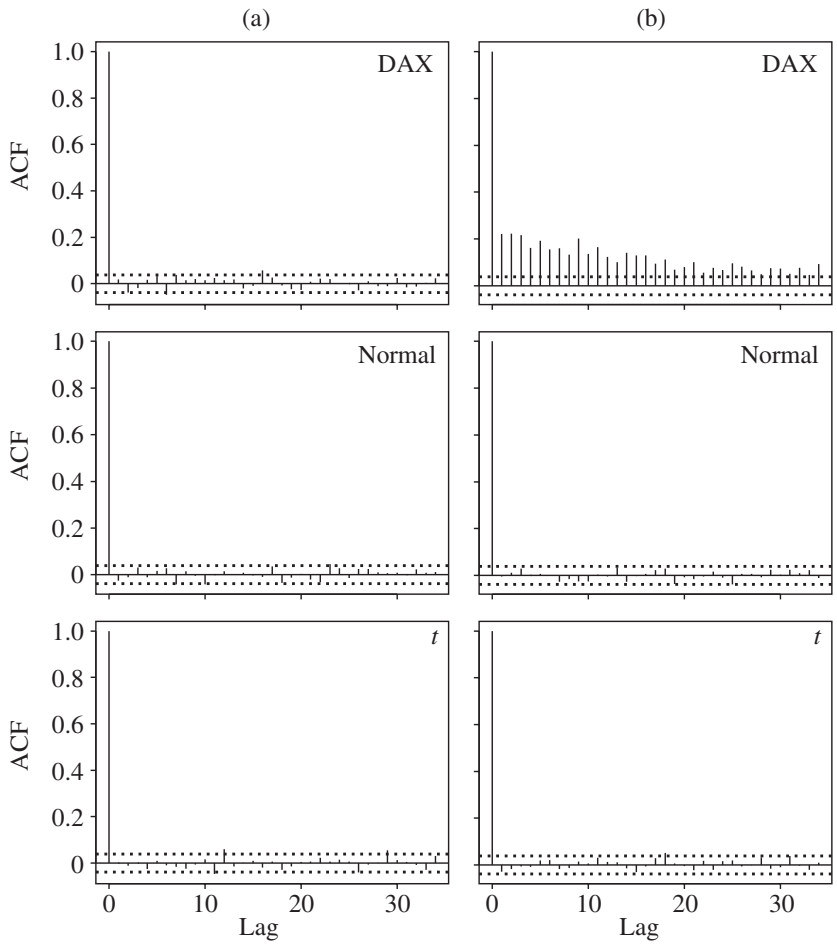


(a) Log-returns for the DAX index from 1985-01-02 to 1994-12-30 ($n = 2608$).

(b) Simulated iid data from a fitted normal with $\hat{\mu} = \bar{X}_n$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ show too few extremes.

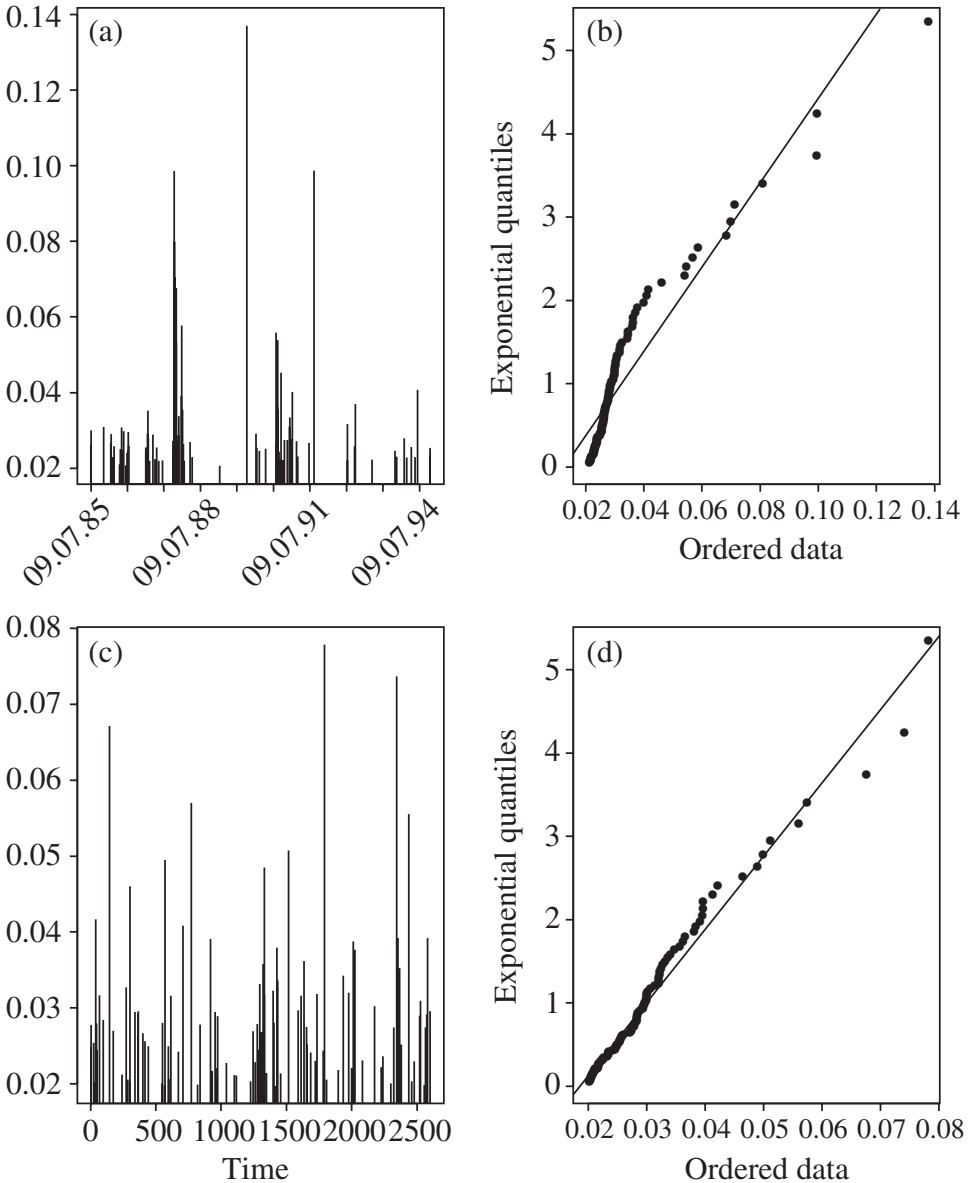
(c) Simulated iid data from a fitted $t_{3.8}$. Better range of values but still no volatility clustering (= tendency for extreme returns to be followed by extreme returns).

Estimated *autocorrelation function (ACF)* $\rho(h) = \text{corr}(X_0, X_h)$, $h \in \mathbb{Z}$



- (a) Sample ACF of $(X_t)_{t\in\mathbb{Z}}$
  (b) and $(|X_t|)_{t\in\mathbb{Z}}$ (similarly: $(X_t^2)_{t\in\mathbb{Z}}$)
- (a): Positive ACF at lag 1 would imply a tendency for returns to be followed by returns of equal sign $\Rightarrow$ not the case $\Rightarrow$ predicted returns ≈ 0
- (b): Positive ACF at lag 1 would imply a tendency for large (small) returns to be followed by large (small) returns $\Rightarrow$ the case for the DAX data $\Rightarrow$ returns cluster (not iid).
- $(X_t)_{t\in\mathbb{Z}}$ not iid $\Rightarrow$ $(Z_t)_{t\in\mathbb{Z}}$ not a Brownian motion $\Rightarrow$ $(S_t)_{t\in\mathbb{Z}}$ not a GBM.

Concerning clustering of extremes, consider the 100 largest losses of the...



- (a) ... DAX index
  (c) ... simulated fitted $t_{3.8}$
- (b), (d) Q-Q plots of waiting times between these large losses (should be $\mathrm{Exp}(\lambda)$ for iid data);
- The DAX data shows longer and shorter waiting times than the iid data, so clustering of extremes.

3.1.2 Non-normality and heavy tails

Formal statistical tests of normality

- For general univariate df F:
  - Kolmogorov–Smirnov (test statistic $T_n = \sup_x |\hat{F}_n(x) - F(x)|$)
  - Cramér–von Mises ($T_n = n \int_{-\infty}^{\infty} (\hat{F}_n(x) - F(x))^2 \, \mathrm{d}F(x)$)
  - Anderson–Darling ($T_n = n \int_{-\infty}^{\infty} \frac{(\hat{F}_n(x) - F(x))^2}{F(x)(1-F(x))} \, \mathrm{d}F(x)$; recommended by D'Agostino and Stephens (1986))
- For $F = \mathrm{N}(\mu, \sigma^2)$:
  - Shapiro–Wilk (idea: quantify Q-Q plot in one number, biased by n)
  - D'Agostino (based on skewness and kurtosis, as Jarque–Bera)
  - Jarque–Bera test: Compares skewness $\beta = \frac{\mathbb{E}((X-\mu)^3)}{\sigma^3}$ and kurtosis $\kappa = \frac{\mathbb{E}((X-\mu)^4)}{\sigma^4}$ with sample versions. The test statistic is $T_n = \frac{n}{6}\left(\hat{\beta}^2 + \frac{1}{4}(\hat{\kappa} - 3)^2\right) \underset{n \text{ large}}{\overset{H_0}{\sim}} \chi^2_2$.

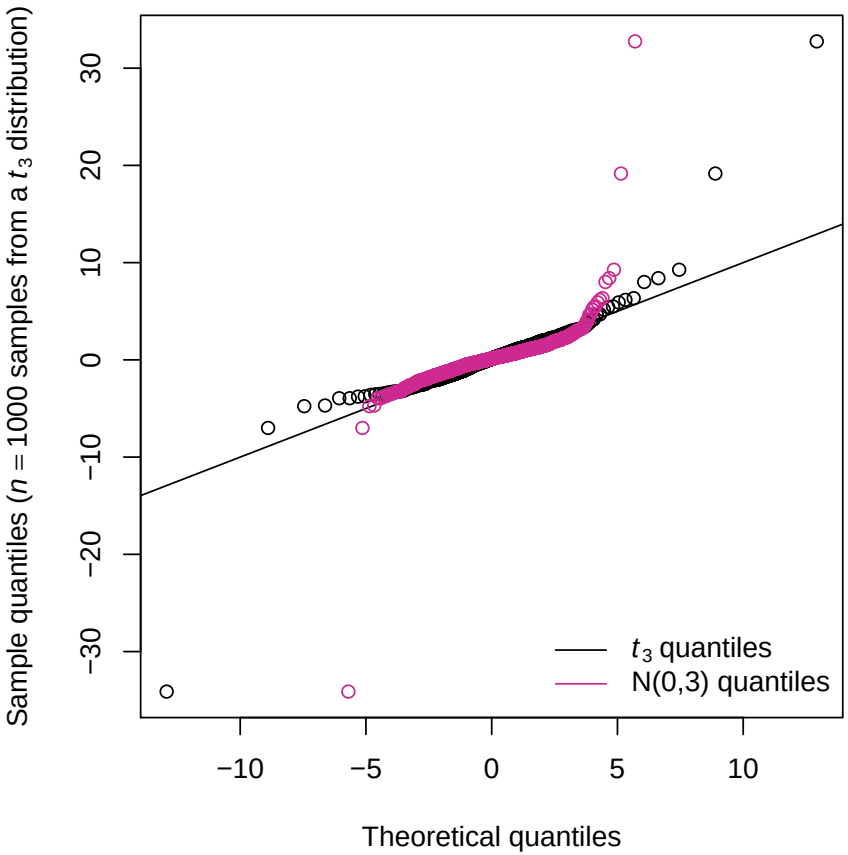
Graphical tests

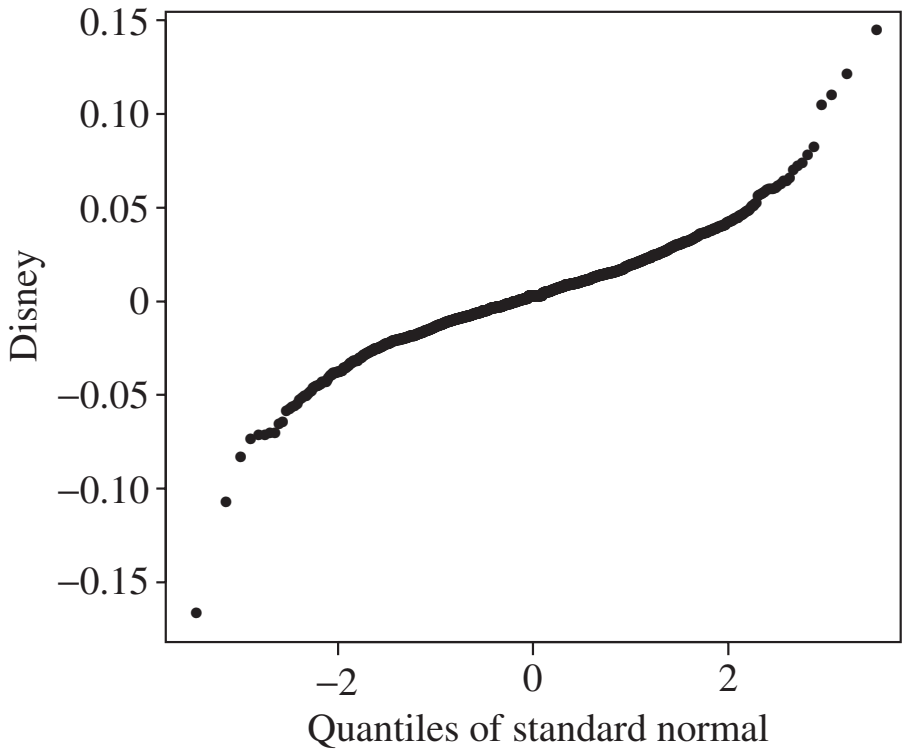
- We can also graphically test whether $X_1, \dots, X_n \sim F$ for some df F based on realizations of iid $X_1, \dots, X_n$.
- Let $X_{(1)} \leq \cdots \leq X_{(n)}$ denote the corresponding order statistics and note that

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq x\}} = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_{(i)} \leq x\}}, \quad x \in \mathbb{R},$$

  i.e. the order statistics contain all relevant information about $X_1, \dots, X_n$.
- Possible graphical tests:
  - P-P plot: For $p_i \underset{n>10}{=} \frac{i-1/2}{n} \approx \frac{i}{n}$, plot $\{(p_i, F(X_{(i)})) : i = 1, \dots, n\}$. If $F \approx \hat{F}_n$, $F(X_{(i)}) \approx p_i$, so the points lie on a line with slope 1.
  - Q-Q plot: Plot $\{(F^{\leftarrow}(p_i), X_{(i)}) : i = 1, \dots, n\}$ (tail differences better visible).

Interpreting Q-Q plots (S-shape hints at a leptokurtic distribution, i.e., narrower center, heavier tails than $\mathrm{N}(\mu, \sigma^2)$ (kurtosis $\kappa = 3$)):





Daily returns typically have kurtosis $\kappa > 3$.

To summarize, we can infer the following stylized facts about univariate financial return series:

(U1) Return series are not iid although they show little serial correlation;

(U2) Series of absolute or squared returns show profound serial correlation;

(U3) Conditional expected returns are close to zero;

(U4) Volatility (conditional standard deviation) appears to vary over time;

(U5) Extreme returns appear in clusters;

(U6) Return series are leptokurtic or heavy-tailed (power-like tails).

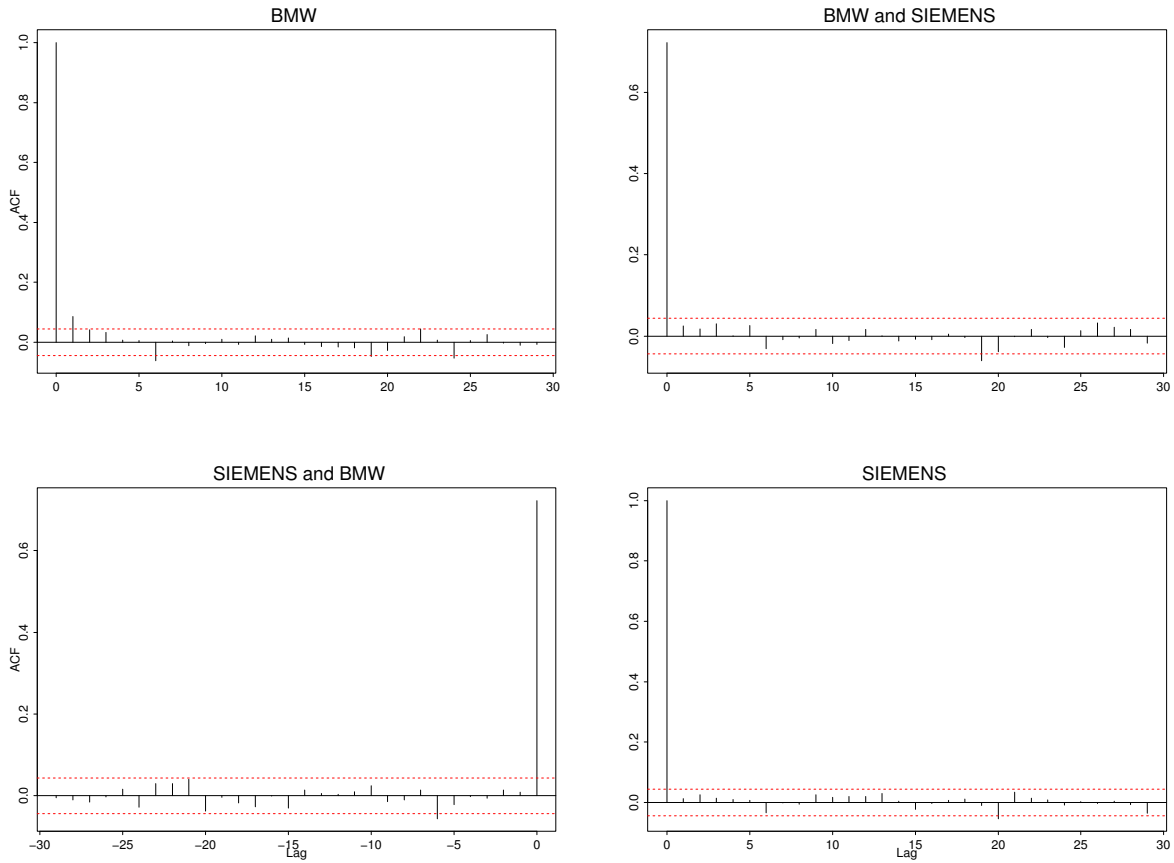
3.2 Multivariate stylized facts

Consider multivariate (componentwise) log-return data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$.

3.2.1 Correlation between series

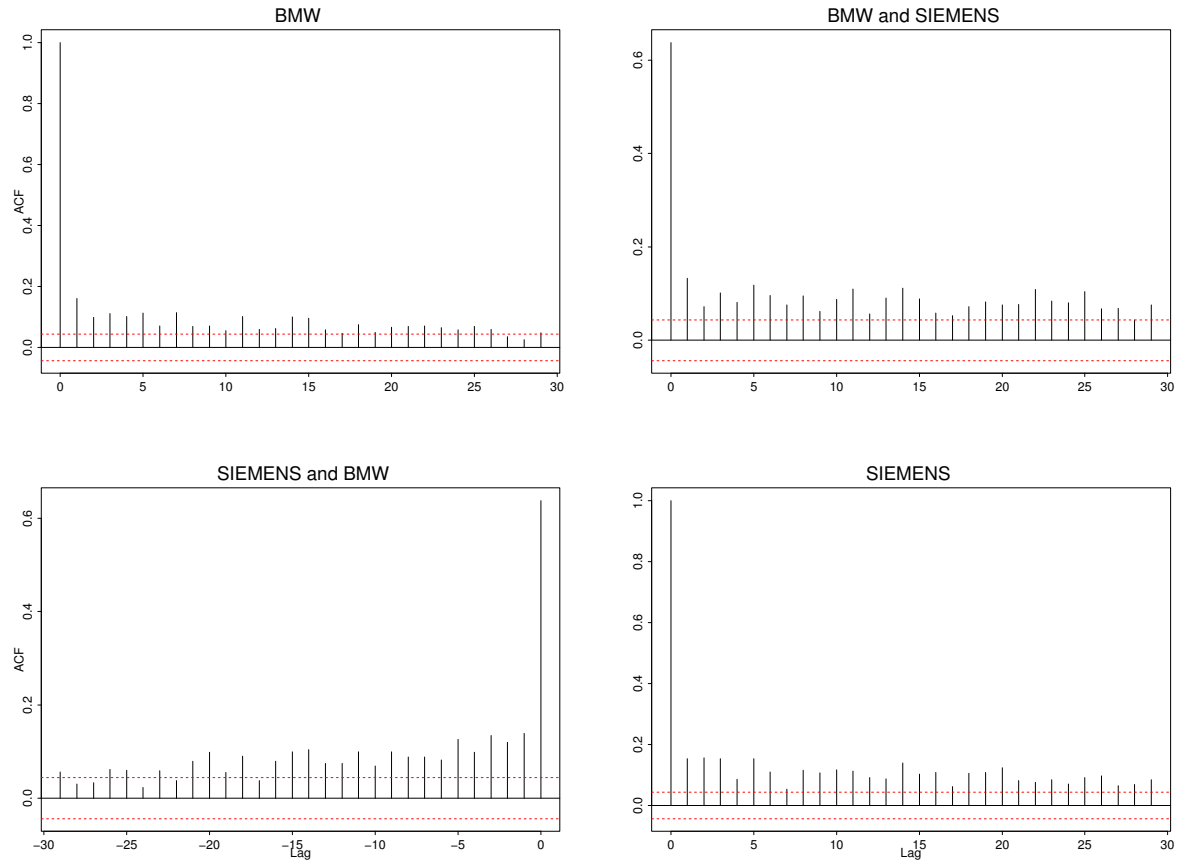
- By (U1), the returns of stock A at t and $t+h$ show little (auto)correlation. The returns of stock A at t and stock B at $t+h$, $h > 0$, also show little cross-correlation. However, Stock A and stock B on day t may be correlated due to factors that affect the whole market (*contemporaneous dependence*).
- These correlations of returns at t vary over time (difficult to detect whether changes are continual or constant within regimes; fit different models for changing correlation, then make a formal comparison).
- Periods of high/low volatility are typically common to more than one stock, so returns of large magnitude in A at t may be followed by returns of large magnitude in A and B at $t+h$.

Estimated correlations between/within series:



Based on 2000 values from period 1985-01-23 to 1994-09-22. Little autocorrelation, little crosscorrelation (at different lags), contemporaneous correlation.

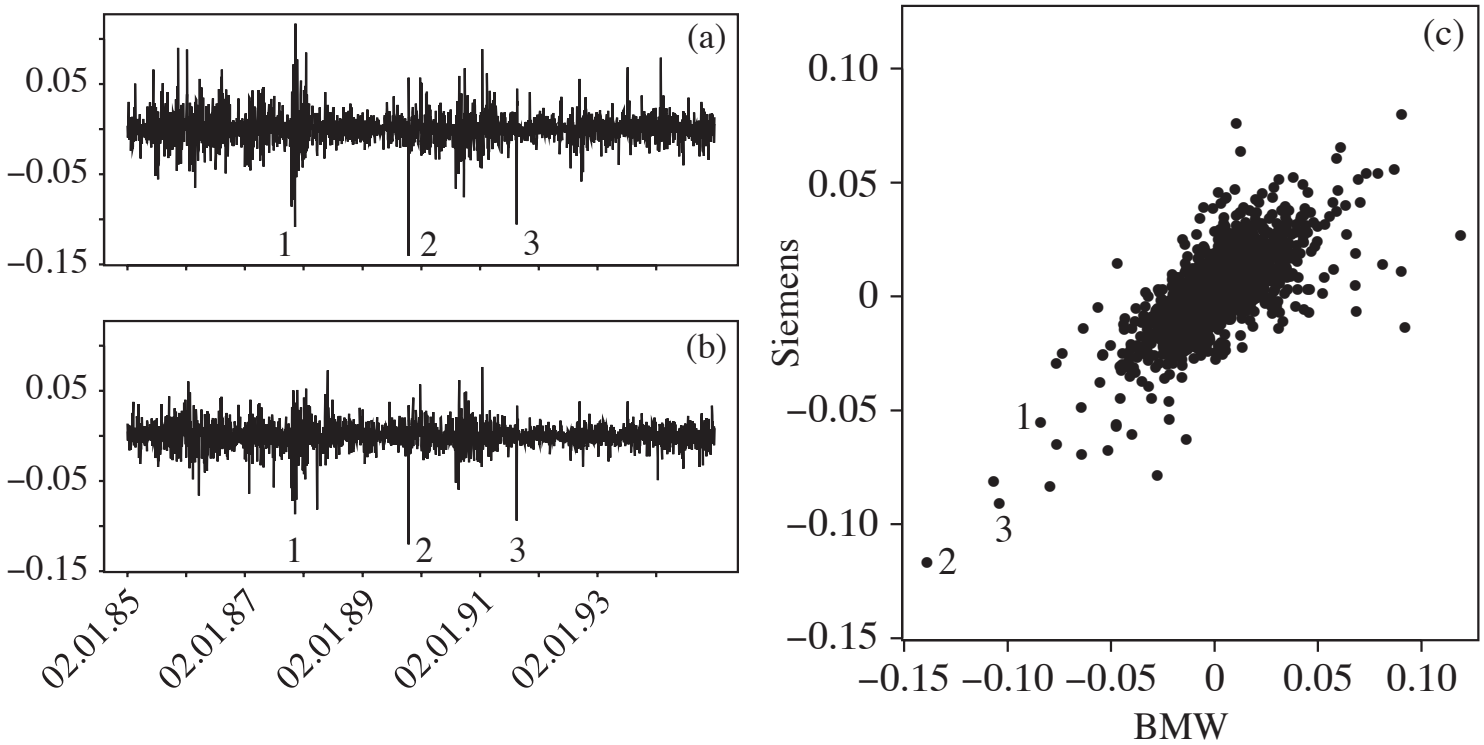
Estimated correlations between/within series of absolute values:



Autocorrelation of absolute returns (indication of volatility clustering).
Common to more than one stock.

3.2.2 **Tail dependence** (for quantifying joint extremes)

(BMW, Siemens) log-returns from 1985-01-23 to 1994-09-22 ($n = 2000$)



In volatile/extreme(ly bad) periods, dependence seems stronger (1: 1987-10-19 Black Monday (DJ drop by 22%); 2: 1989-10-16 Monday demonstrations (Wende); 3: 1991-08-19 coup against soviet president M. Gorbachev).

To summarize, we can infer the following stylized facts about multivariate financial return series:

(M1) Multivariate return series show little evidence of cross-correlation, except for contemporaneous returns (i.e. at the same t);

(M2) Multivariate series of absolute returns show profound cross-correlation;

(M3) Correlations between contemporaneous returns vary over time (difficult to infer from empirical correlations due to estimation error in small samples);

(M4) Extreme returns in one series often coincide with extreme returns in several other series (e.g. tail dependence).

4 Financial time series

4.1 Fundamentals of time series analysis

4.2 GARCH models for changing volatility

4.1 Fundamentals of time series analysis

4.1.1 Basic definitions

A *stochastic process* is a family of rvs $(X_t)_{t\in I}$, $I \subseteq \mathbb{R}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A *time series* is a discrete-time ($I \subseteq \mathbb{Z}$) stochastic process.

Definition 4.1 (Mean function, autocovariance function)
Assuming they exist, the *mean function* $\mu(t)$ and the *autocovariance function* $\gamma(t,s)$ of $(X_t)_{t\in\mathbb{Z}}$ are defined by

$$\mu(t) = \mathbb{E}(X_t), \quad t \in \mathbb{Z},$$
$$\gamma(t,s) = \mathrm{cov}(X_t, X_s) = \mathbb{E}((X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)), \quad t, s, \in \mathbb{Z}.$$

Definition 4.2 ((Weak/strict) stationarity)

1) $(X_t)_{t\in\mathbb{Z}}$ is *(weakly/covariance) stationary* if $\mathbb{E}(X_t^2) < \infty$, $\mu(t) = \mu \in \mathbb{R}$ and $\gamma(t,s) = \gamma(t+h, s+h)$ for all $t, s, h \in \mathbb{Z}$.
2) $(X_t)_{t\in\mathbb{Z}}$ is *strictly stationary* if $(X_{t_1}, \ldots, X_{t_n}) \overset{\text{d}}{=} (X_{t_1+h}, \ldots, X_{t_n+h})$ for all $t_1, \ldots, t_n, h \in \mathbb{Z}$, $n \in \mathbb{N}$.

Remark 4.3

1) Both types of stationarity formalize the idea that $(X_t)_{t\in\mathbb{Z}}$ behaves similarly in any time period.
2) - Strict stationarity $\not\Rightarrow$ stationarity (unless also $\mathbb{E}(X_t^2)$ exists).
   - Stationarity $\not\Rightarrow$ strict stationarity ($\mathbb{E}(|X_t|^p)$, $p > 2$, could change).
3) If $(X_t)_{t\in\mathbb{Z}}$ is stationary, $\gamma(0, t-s) = \gamma(s,t) = \gamma(t,s) = \gamma(0, s-t)$, so $\gamma(t,s)$ only depends on the *lag* $h = |t-s|$. We can thus define $\gamma(h) := \gamma(0, |h|)$, $h \in \mathbb{Z}$.

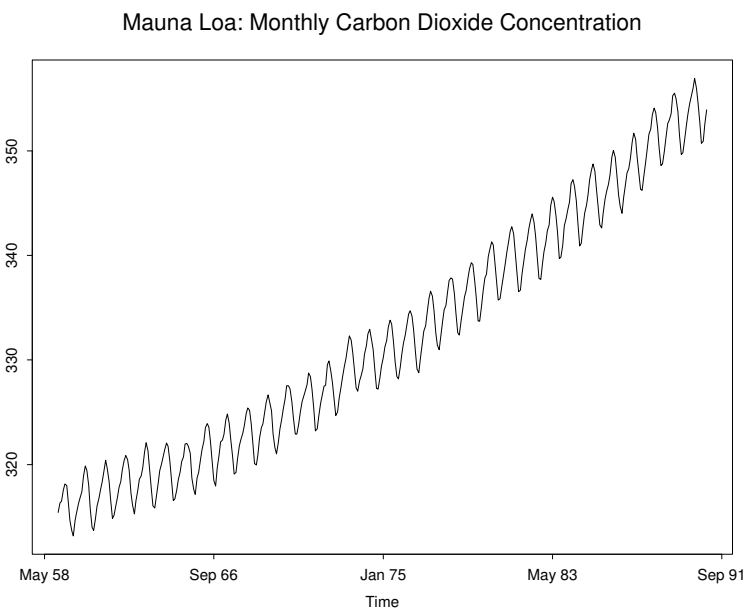
Autocorrelation in stationary time series

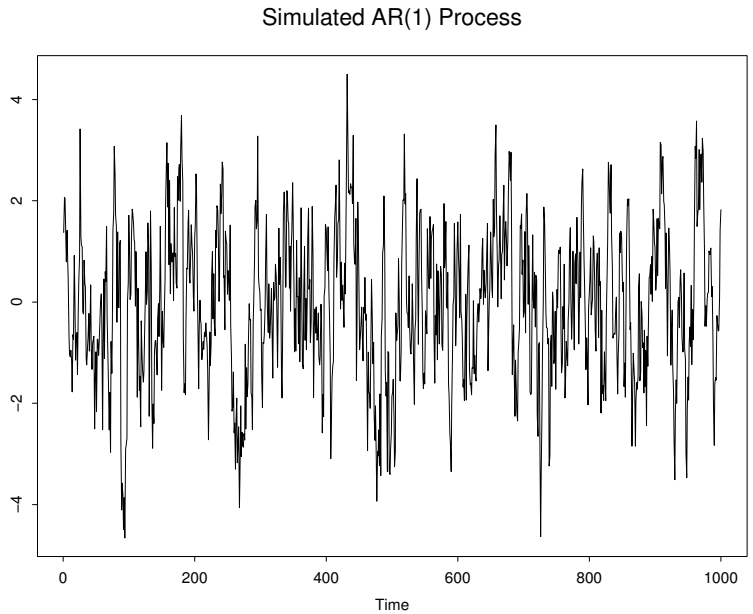
Definition 4.4 (ACF)

The *autocorrelation function (ACF)* (or *serial correlation*) of a stationary time series $(X_t)_{t\in\mathbb{Z}}$ is defined by

$$\rho(h) := \operatorname{corr}(X_0, X_h) = \gamma(h)/\gamma(0), \quad h \in \mathbb{Z}.$$

Stationary?





White noise processes

Definition 4.5 ((Strict) white noise)

1) $(X_t)_{t\in\mathbb{Z}}$ is a *white noise* process if $(X_t)_{t\in\mathbb{Z}}$ is stationary with $\rho(h) = I_{\{h=0\}}$ (no serial correlation). If $\mu(t) = 0$, $\gamma(0) = \mathrm{var}(X_t) = \sigma^2$, $(X_t)_{t\in\mathbb{Z}}$ is denoted by $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$.
2) $(X_t)_{t\in\mathbb{Z}}$ is a *strict white noise* process if $(X_t)_{t\in\mathbb{Z}}$ is an iid sequence of rvs with $\gamma(0) = \mathrm{var}(X_t) = \sigma^2 < \infty$. If $\mu(t) = 0$, we write $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{SWN}(0, \sigma^2)$.

Definition 4.6 (MGDS)

$(X_t)_{t\in\mathbb{Z}}$ is a *martingale-difference sequence (MGDS)* w.r.t. $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ (typically the natural filtration $\mathcal{F}_t = \sigma(\{X_s : s \le t\})$) if $\mathbb{E}|X_t| < \infty$, $t \in \mathbb{Z}$, $(X_t)_{t\in\mathbb{Z}}$ is adapted to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ (i.e. $X_t \in \mathcal{F}_t$, $t \in \mathbb{Z}$); and $\mathbb{E}(X_{t+1} \,|\, \mathcal{F}_t) = 0$ for all $t \in \mathbb{Z}$.

4.1.2 ARMA processes

Definition 4.7 ($\mathrm{ARMA}(p,q)$)
Let $(\varepsilon_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0,\sigma^2)$. $(X_t)_{t\in\mathbb{Z}}$ is a *zero-mean* $\mathrm{ARMA}(p,q)$ *process* if it is stationary and satisfies, for all $t \in \mathbb{Z}$,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}. \quad (6)$$

$(X_t)_{t\in\mathbb{Z}}$ is $\mathrm{ARMA}(p,q)$ with *mean* μ if $(X_t - \mu)_{t\in\mathbb{Z}}$ is a zero-mean $\mathrm{ARMA}(p,q)$.

- The defining equation (6) can be written as $\phi(B)X_t = \theta(B)\varepsilon_t$, $t \in \mathbb{Z}$, where B denotes the *backshift operator* (such that $B^k X_t = X_{t-k}$) and $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$.
- For practical purposes, it suffices to consider *causal* ARMA processes $(X_t)_{t\in\mathbb{Z}}$ satisfying $X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}$ (only depending on past) for $\sum_{k=0}^{\infty} |\psi_k| < \infty$ (*absolute summability condition*; guarantees $\mathbb{E}|X_t| < \infty$).

Theorem 4.8 (Stationary and causal ARMA solutions)
Let $(X_t)_{t\in\mathbb{Z}}$ be an $\mathrm{ARMA}(p,q)$ process for which $\phi(z), \theta(z)$ have no roots in common. Then

$$(X_t)_{t\in\mathbb{Z}} \text{ is stationary and causal} \quad \Leftrightarrow \quad \phi(z) \neq 0 \quad \forall\, z \in \mathbb{C} : |z| \leq 1.$$

In this case, $X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}$ for $\sum_{k=0}^{\infty} \psi_k z^k = \theta(z)/\phi(z)$, $|z| \leq 1$.

- If $\theta(z) \neq 0$, $|z| \leq 1$ (known as *invertibility condition*), $(X_t)_{t\in\mathbb{Z}}$ is *invertible*, i.e. we can recover ε_t from $(X_s)_{s\leq t}$ (via $\varepsilon_t = \phi(B)X_t/\theta(B)$), so $\varepsilon_t \in \mathcal{F}_t = \sigma(\{X_s : s \leq t\})$.
- An $\mathrm{ARMA}(p,q)$ process with mean μ can be written as $X_t = \mu_t + \varepsilon_t$ for $\mu_t = \mu + \sum_{k=1}^{p} \phi_k(X_{t-k} - \mu) + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k}$. If $(X_t)_{t\in\mathbb{Z}}$ is invertible, $\mu_t \in \mathcal{F}_{t-1}$. If $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a MGDS w.r.t. $(\mathcal{F}_t)_{t\in\mathbb{Z}}$, then $\mu_t = \mathbb{E}(X_t \,|\, \mathcal{F}_{t-1})$. Therefore, ARMA processes put structure on the conditional mean μ_t given the past. We will see that GARCH processes put structure on $\sigma_t^2 = \mathrm{var}(X_t \,|\, \mathcal{F}_{t-1})$ (helpful for modeling volatility clustering).

4.1.3 Analysis in the time domain

Correlogram

A *correlogram* is a plot of $(h, \hat{\rho}(h))_{h \geq 0}$ for the sample ACF

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{n}(X_{t+h} - \bar{X}_n)(X_t - \bar{X}_n)}{\sum_{t=1}^{n}(X_t - \bar{X}_n)^2}, \quad h \in \{0, \ldots, n\}.$$

Theorem 4.9

Let $X_t - \mu = \sum_{k=0}^{\infty} \psi_k Z_{t-k}$ and $(Z_t) \sim \mathrm{SWN}(0, \sigma^2)$. Under suitable conditions,

$$\sqrt{n}\left(\begin{pmatrix}\hat{\rho}(1)\\ \vdots \\ \hat{\rho}(h)\end{pmatrix} - \begin{pmatrix}\rho(1)\\ \vdots \\ \rho(h)\end{pmatrix}\right) \underset{(n\to\infty)}{\overset{\mathsf{d}}{\to}} \mathrm{N}_h(\mathbf{0}, W), \quad h \in \mathbb{N},$$

for a matrix W depending on ρ; see MFE (2015, Theorem 4.13).

If the ARMA process is SWN itself, $\sqrt{n}\begin{pmatrix}\hat{\rho}(1)\\ \vdots \\ \hat{\rho}(h)\end{pmatrix} \underset{(n\to\infty)}{\xrightarrow{\mathsf{d}}} \mathrm{N}_h(\mathbf{0}, I_h)$, so that with probability $1-\alpha$,

$$\hat{\rho}(k) \underset{(n \text{ large})}{\in} \left[-\frac{q_{1-\alpha/2}}{\sqrt{n}}, \frac{q_{1-\alpha/2}}{\sqrt{n}}\right], \quad k \in \{1, \dots, h\},$$

where $q_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$. This interval is usually shown in correlogram.

- As a formal test of the SWN hypothesis, one can use the Ljung–Box test with test statistic

$$T = n(n+2)\sum_{k=1}^{h} \frac{\hat{\rho}(k)^2}{n-k} \underset{n \text{ large}}{\sim} \chi^2_h; \quad \text{reject if } T > \chi^{2\,-1}_h(1-\alpha).$$

- If $(X_t)_{t\in\mathbb{Z}}$ is SWN, so is $(X_t^2)_{t\in\mathbb{Z}}$. It is a good idea to also apply the correlogram and Ljung–Box tests to $(|X_t|)_{t\in\mathbb{Z}}$ or $(X_t^2)_{t\in\mathbb{Z}}$.

4.2 GARCH models for changing volatility

- (G)ARCH = (generalized) autoregressive conditionally heteroscedastic
- They are the most important models for daily risk-factor returns.

Definition 4.10 ($\mathrm{GARCH}(p,q)$)
Let $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{SWN}(0,1)$ (typically $Z_t \overset{\text{ind.}}{\sim} \mathrm{N}(0,1)$ or $Z_t \overset{\text{ind.}}{\sim} t_\nu(0,(\nu-2)/\nu)$). $(X_t)_{t\in\mathbb{Z}}$ is a $\mathrm{GARCH}(p,q)$ *process* if it is strictly stationary and satisfies

$$X_t = \sigma_t Z_t,$$

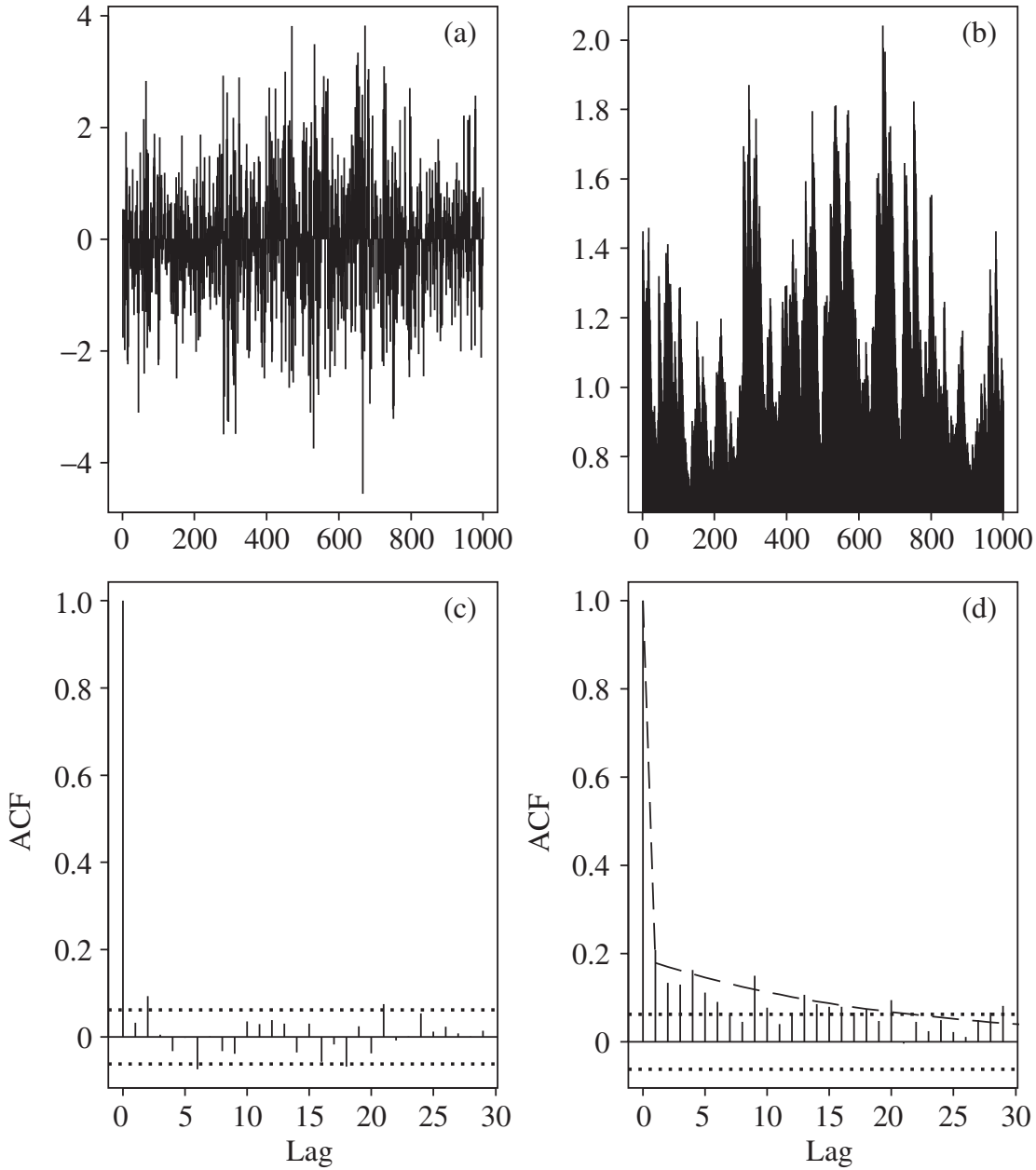
$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k X_{t-k}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2,$$

where $\alpha_0 > 0$, $\alpha_k \geq 0$, $k \in \{1,\dots,p\}$, $\beta_k \geq 0$, $k \in \{1,\dots,q\}$.

If one of $|X_{t-1}|,\dots,|X_{t-p}|$ or $\sigma_{t-1},\dots,\sigma_{t-q}$ is large, X_t is drawn from a distribution with (persistently) large variance.

Example 4.11 ($\mathrm{GARCH}(1,1)$)

- One can show (via stoch. recurrence relations) that a $\mathrm{GARCH}(1,1)$ process $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary if $\mathbb{E}(\log(\alpha_1 Z_t^2+\beta_1))<\infty$. In this case, $X_t = Z_t\sqrt{\alpha_0(1+\sum_{k=1}^{\infty}\prod_{j=1}^{k}(\alpha_1 Z_{t-j}^2+\beta_1))}$.
- $(X_t)_{t\in\mathbb{Z}}$ is stationary $\Leftrightarrow \alpha_1+\beta_1<1$. In this case, $\mathrm{var}(X_t)=\frac{\alpha_0}{1-\alpha_1-\beta_1}$.
- $\mathrm{GARCH}(1,1)$ is typically leptokurtic:
  Provided that $\mathbb{E}((\alpha_1 Z_t^2+\beta_1)^2)<1$ (or $(\alpha_1+\beta_1)^2<1-(\kappa(Z_t)-1)\alpha_1^2$), one can show that $\kappa(X_t)=\frac{\kappa(Z_t)(1-(\alpha_1+\beta_1)^2)}{1-(\alpha_1+\beta_1)^2-(\kappa(Z_t)-1)\alpha_1^2}$.
  If $\kappa(Z_t)>1$ (Gaussian, scaled t innovations), $\kappa(X_t)>\kappa(Z_t)$.
- Parallels with the ARMA(1,1) process:
  If $\mathbb{E}(X_t^4)<\infty$, $\alpha_1+\beta_1<1$ and $\varepsilon_t=\sigma_t^2(Z_t^2-1)$, one can show that $(X_t^2)_{t\in\mathbb{Z}}$ is an $\mathrm{ARMA}(1,1)$ of the form $X_t^2-\frac{\alpha_0}{1-\alpha_1-\beta_1}=(\alpha_1+\beta_1)(X_{t-1}^2-\frac{\alpha_0}{1-\alpha_1-\beta_1})+\varepsilon_t-\beta_1\varepsilon_{t-1}$.



a) $n = 1000$ realization of a GARCH$(1,1)$ process with $\alpha_0 = 0.5$, $\alpha_1 = 0.1$, $\beta_1 = 0.85$ and Gaussian innovations;

b) Realization of the volatility $(\sigma_t)_{t\in\mathbb{Z}}$;

c) Correlogram of $(X_t)_{t\in\mathbb{Z}}$, can be shown to be $I_{\{h=0\}}$

d) Correlogram of $(X_t^2)_{t\in\mathbb{Z}}$ (ARMA(1,1)); dashed line = true ACF

4.2.1 Simple extensions of the GARCH model

Consider stationary GARCH processes as white noise for ARMA processes.

Definition 4.12 ($\mathrm{ARMA}(p_1, q_1)$ with $\mathrm{GARCH}(p_2, q_2)$ errors)
Let $(Z_t)_{t\in\mathbb{Z}} \sim \mathrm{SWN}(0,1)$. $(X_t)_{t\in\mathbb{Z}}$ is an $\mathrm{ARMA}(p_1, q_1)$ *process with* $\mathrm{GARCH}(p_2, q_2)$ *errors* if it is stationary and satisfies

$$X_t = \mu_t + \varepsilon_t \quad \text{for} \quad \varepsilon_t = \sigma_t Z_t \quad (\text{so } X_t = \mu_t + \sigma_t Z_t),$$

$$\mu_t = \mu + \sum_{k=1}^{p_1} \phi_k (X_{t-k} - \mu) + \sum_{k=1}^{q_1} \theta_k (\underbrace{X_{t-k} - \mu_{t-k}}_{=\varepsilon_{t-k}}),$$

$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^{p_2} \alpha_k (X_{t-k} - \mu_{t-k})^2 + \sum_{k=1}^{q_2} \beta_k \sigma_{t-k}^2,$$

where $\alpha_0 > 0$, $\alpha_k \geq 0$, $k \in \{1, \dots, p_2\}$, $\beta_k \geq 0$, $k \in \{1, \dots, q_2\}$, $\sum_{k=1}^{p_2} \alpha_k + \sum_{k=1}^{q_2} \beta_k < 1$.

4.2.2 Fitting GARCH models to data

- The most widely used approach is maximum likelihood.
- After model fitting, we check residuals. Consider an ARMA model with GARCH errors $X_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t Z_t$.
- The standardized residuals

$$\hat{Z}_t = \hat{\varepsilon}_t / \hat{\sigma}_t, \quad \hat{\sigma}_t^2 = \hat{\alpha}_0 + \sum_{k=1}^{p_2} \hat{\alpha}_k \hat{\varepsilon}_{t-k}^2 + \sum_{k=1}^{q_2} \hat{\beta}_k \hat{\sigma}_{t-k}^2; \tag{7}$$

  (with starting values for $\hat{\varepsilon}_t$ as 0 and $\hat{\sigma}_t$ as the sample variances) should behave like SWN. Check this via correlograms of $(\hat{Z}_t)$ and $(|\hat{Z}_t|)$ and by applying the Ljung–Box test of strict white noise.
- In case of no rejection (the dynamics have been satisfactorily captured), the validity of the innovation distribution can also be assessed (e.g. via Q-Q plots or goodness-of-fit tests).

Forecasting VaR_α and ES_α

- Suppose we now want to forecast $\mathrm{VaR}_\alpha^{t+1}$, ES_α^{t+1}, risk measures based on $F_{X_{t+1}|\mathcal{F}_t}$. If $Z_t \overset{\text{ind.}}{\sim} F_Z$, the $\mathcal{F}_t$-measurability of μ_{t+1} and σ_{t+1}, and $X_{t+1} = \mu_{t+1} + \sigma_{t+1} Z_{t+1}$ imply that

$$F_{X_{t+1}|\mathcal{F}_t}(x) = \mathbb{P}(\mu_{t+1} + \sigma_{t+1} Z_{t+1} \leq x \,|\, \mathcal{F}_t) = F_Z\Big(\frac{x - \mu_{t+1}}{\sigma_{t+1}}\Big).$$

- Then $\mathrm{VaR}_\alpha^{t+1} = \mu_{t+1} + \sigma_{t+1} F_Z^{\leftarrow}(\alpha)$ and $\mathrm{ES}_\alpha^{t+1} = \mu_{t+1} + \sigma_{t+1} \mathrm{ES}_\alpha(Z)$.
- If we have estimated σ_{t+1} (and μ_{t+1}; often taken as 0) it remains to estimate $F_Z^{\leftarrow}(\alpha)$ and $\mathrm{ES}_\alpha(Z)$ (easy for GARCH-type models)

5 Extreme value theory

5.1 Maxima

5.2 Threshold exceedances

5.1 Maxima

Consider a series of financial losses $(X_k)_{k\in\mathbb{N}}$.

5.1.1 Generalized extreme value distribution

Convergence of sums

Let $(X_k)_{k\in\mathbb{N}}$ be iid with $\mathbb{E}(X_1^2) < \infty$ (mean μ, variance σ^2) and $S_n = \sum_{k=1}^n X_k$. As $n \to \infty$, $\bar{X}_n \overset{\text{a.s.}}{\to} \mu$ by the Strong Law of Large Numbers (SLLN), so $(\bar{X}_n - \mu)/\sigma \overset{\text{a.s.}}{\to} 0$. By the CLT,

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \underset{n\uparrow\infty}{\overset{\text{d}}{\to}} \mathrm{N}(0,1) \text{ or } \lim_{n\to\infty} \mathbb{P}\Big(\frac{S_n - d_n}{c_n} \le x\Big) = \Phi(x),$$

where the sequences $c_n = \sqrt{n}\sigma$ and $d_n = n\mu$ give normalization and where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2}\,\mathrm{d}z$.

Convergence of maxima

QRM is concerned with maximal losses (worst-case losses). Let $(X_i)_{i\in\mathbb{N}} \overset{\text{ind.}}{\sim} F$ (can be relaxed to a strictly stationary time series) and F continuous. Then the *block maximum* is given by

$$M_n = \max\{X_1, \dots, X_n\}.$$

One can show that $M_n \overset{\text{a.s.}}{\underset{n\to\infty}{\to}} x_F$ (similar as in the SLLN; due to monotone convergence to a constant) where

$$x_F := \sup\{x \in \mathbb{R} : F(x) < 1\} = F^{\leftarrow}(1) \leq \infty$$

denotes the *right endpoint of* F.

Question: Is there a "CLT" for block maxima?

Idea CLT: What about linear transformations (the simplest possible)?

Definition 5.1 (Maximum domain of attraction)
Suppose we find *normalizing sequences* of real numbers $(c_n) > 0$ and (d_n) such that $(M_n - d_n)/c_n$ converges in distribution, i.e.

$$\begin{aligned}\mathbb{P}((M_n - d_n)/c_n \leq x) &= \mathbb{P}(M_n \leq c_n x + d_n) \\ &= \mathbb{P}(X_i \leq c_n x + d_n,\ i = 1, \ldots, n) \\ &= F^n(c_n x + d_n) \underset{n \uparrow \infty}{\to} H(x),\end{aligned}$$

for some *non-degenerate* df H (not a unit jump). Then F is in the *maximum domain of attraction* of H ($F \in \mathrm{MDA}(H)$).

The convergence to types theorem guarantees that H is determined up to location/scale, i.e. H specifies a unique *type* of distribution.
Question: What does H look like?

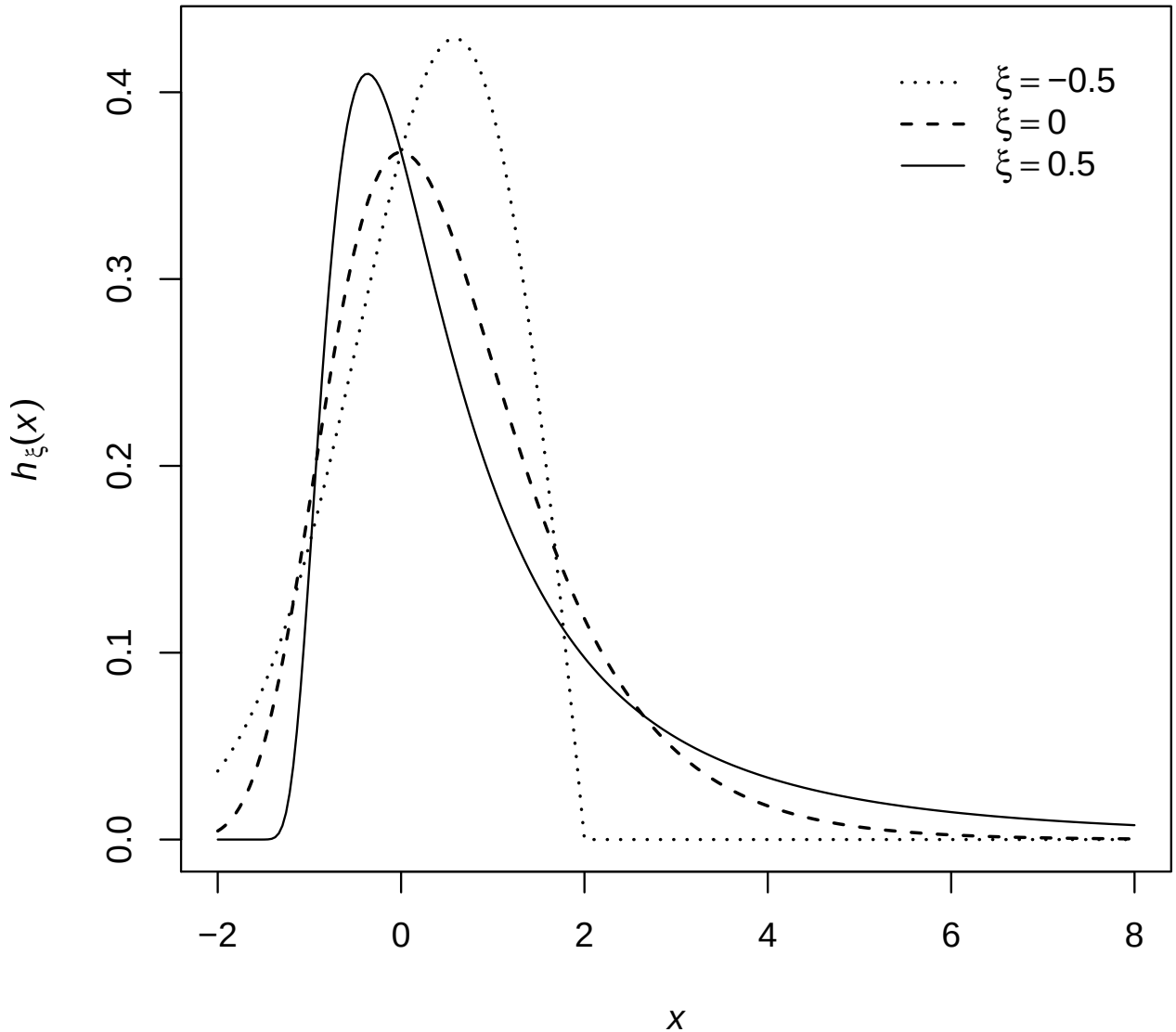
Definition 5.2 (Generalized extreme value (GEV) distribution)
The (standard) *generalized extreme value (GEV) distribution* is given by

$$H_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}), & \text{if } \xi \neq 0, \\ \exp(-e^{-x}), & \text{if } \xi = 0, \end{cases}$$

where $1 + \xi x > 0$ (MLE!). A three-parameter family is obtained by a location-scale transform $H_{\xi,\mu,\sigma}(x) = H_\xi((x-\mu)/\sigma)$, $\mu \in \mathbb{R}$, $\sigma > 0$.

- The larger ξ, the heavier tailed H_ξ (if $\xi > 0$, $\mathbb{E}(X^k) = \infty$ iff $k \geq \frac{1}{\xi}$).
- ξ is the *shape* (determines moments, tail). Special cases:
  1) $\xi < 0$: the Weibull df, short-tailed, $x_{H_\xi} < \infty$;
  2) $\xi = 0$: the Gumbel df, $x_{H_0} = \infty$, decays exponentially;
  3) $\xi > 0$: the Fréchet df, $x_{H_\xi} = \infty$, heavy-tailed ($\bar{H}_\xi(x) \approx (\xi x)^{-1/\xi}$), most important case for practice

Density h_ξ for $\xi \in \{-0.5, 0, 0.5\}$ (dotted, dashed, solid)



Theorem 5.3 (Fisher–Tippett–Gnedenko)
If $F \in \mathrm{MDA}(H)$ for some non-degenerate H, then H must be of GEV type, i.e. $H = H_{\xi,\mu,\sigma}$ for some $\xi \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma > 0$.

Proof. Non-trivial. For a sketch, see Embrechts, Klüppelberg, et al. (1997, p. 122). □

- **Interpretation:** If location-scale transformed maxima of iid random variables converge in distribution to a non-degenerate limit, the limiting distribution must be a location-scale transformed GEV distribution (that is, of GEV type).
- One can always choose normalizing sequences $(c_n) > 0$, (d_n) such that $H_{\xi,\mu,\sigma}$ appears in standard form (from a statistical point of view, $(c_n) > 0$, (d_n) can simply be estimated).
- All commonly encountered continuous distributions are in the MDA of some GEV distribution.

Example 5.4 (Pareto distribution)

For $(X_i)_{i\in\mathbb{N}} \overset{\text{ind.}}{\sim} \text{Par}(\theta,\kappa)$ with $F(x) = 1-(\frac{\kappa}{\kappa+x})^\theta$, $x \geq 0$, $\theta,\kappa > 0$, choosing $c_n = \kappa n^{1/\theta}/\theta$, $d_n = \kappa(n^{1/\theta}-1)$, $F^n(c_n x + d_n)$ equals

$$\Big(1-\Big(\frac{\kappa}{\kappa + x(\kappa n^{1/\theta}/\theta) + (\kappa(n^{1/\theta}-1))}\Big)^\theta\Big)^n$$

$$= \Big(1-\Big(\frac{1}{1+xn^{1/\theta}/\theta + n^{1/\theta}-1}\Big)^\theta\Big)^n = \Big(1-\Big(\frac{1}{n^{1/\theta}(1+x/\theta)}\Big)^\theta\Big)^n$$

$$= \Big(1+\frac{-(1+x/\theta)^{-\theta}}{n}\Big)^n \underset{n\uparrow\infty}{\to} \exp(-(1+x/\theta)^{-\theta}) = H_{1/\theta}(x) \text{ (Fréchet)}$$

Therefore, $F \in \text{MDA}(H_{1/\theta})$.

5.1.2 Maximum domains of attraction

All commonly applied continuous F belong to $\mathrm{MDA}(H_\xi)$ for some $\xi \in \mathbb{R}$ and μ, σ can be estimated. But how can we characterize/determine ξ? All $F \in \mathrm{MDA}(H_\xi)$, $\xi > 0$, allow for a characterization based on:

Definition 5.5 (Slowly/regularly varying functions)

1) A positive, Lebesgue-measurable function L on $(0, \infty)$ is *slowly varying at* ∞ if $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$, $t > 0$. The class of all such functions is denoted by $\mathcal{R}_0$; e.g. $c, \log \in \mathcal{R}_0$.
2) A positive, Lebesgue-measurable function h on $(0, \infty)$ is *regularly varying at* ∞ *with index* $\alpha \in \mathbb{R}$ if $\lim_{x\to\infty} \frac{h(tx)}{h(x)} = t^\alpha$, $t > 0$. The class of all such functions is denoted by $\mathcal{R}_\alpha$; e.g. $x^\alpha L(x) \in \mathcal{R}_\alpha$.

If $\bar{F} \in \mathcal{R}_{-\alpha}$, $\alpha > 0$, the tail of F decays like a power function (Pareto like).

The Fréchet case

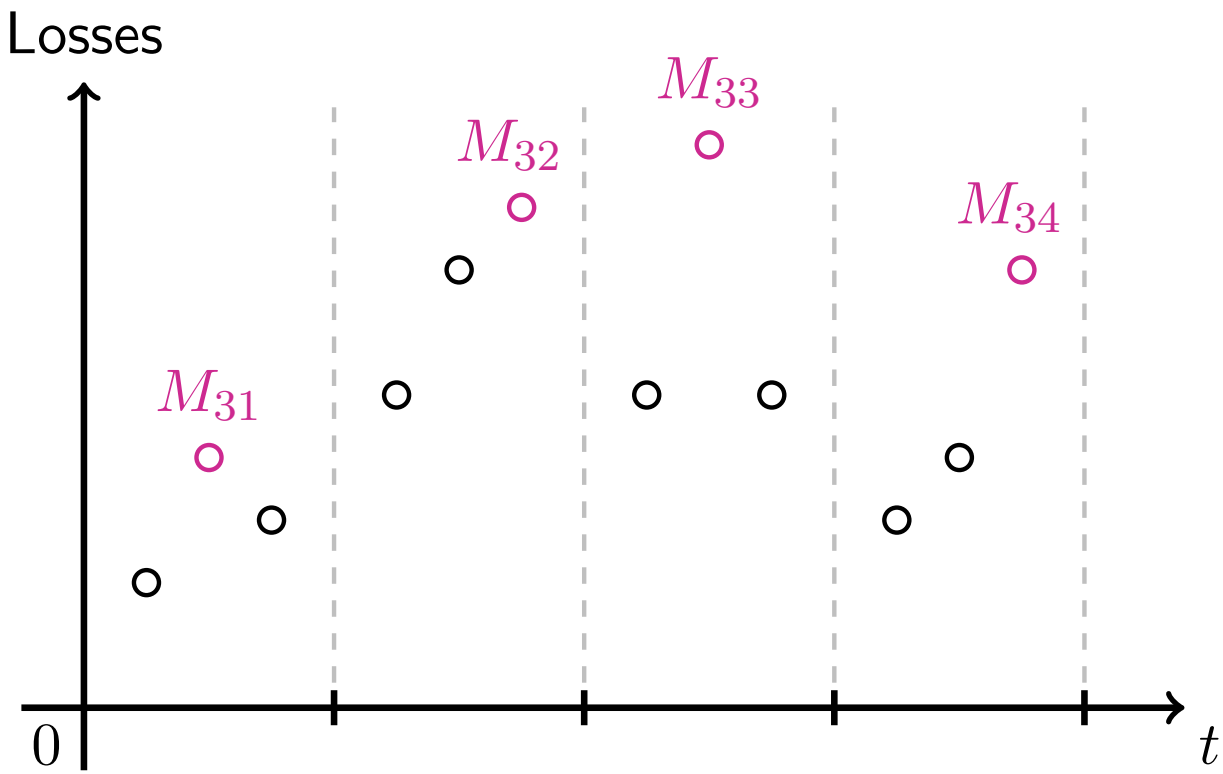
Theorem 5.6 (Fréchet MDA, Gnedenko (1943))
$F \in \mathrm{MDA}(H_\xi)$ for $\xi > 0$ if and only if $\bar{F}(x) = x^{-1/\xi} L(x)$ for some $L \in \mathcal{R}_0$. If $F \in \mathrm{MDA}(H_\xi)$, $\xi > 0$, the normalizing sequences can be chosen as $c_n = F^{\leftarrow}(1 - 1/n)$ and $d_n = 0$, $n \in \mathbb{N}$.

- **Interpretation:** Distributions in $\mathrm{MDA}(H_\xi)$, $\xi > 0$, are those whose tails decay like power functions; $\alpha = 1/\xi$ is known as *tail index*.
- **Examples in $\mathrm{MDA}(H_\xi)$, $\xi > 0$:** Inverse gamma, Student t, log-gamma, F, Cauchy, α-stable with $0 < \alpha < 2$, Burr and Pareto

Example 5.7 (Pareto distribution)
For $F = \mathrm{Par}(\theta, \kappa)$, $\bar{F}(x) = (\kappa/(\kappa + x))^\theta = (1 + x/\kappa)^{-\theta} = x^{-\theta} L(x)$, $x \geq 0$, $\theta, \kappa > 0$, where $L(x) = (x^{-1} + \kappa^{-1})^{-\theta} \in \mathcal{R}_0$. We (again) see that $F \in \mathrm{MDA}(H_\xi)$, $\xi > 0$.

5.1.3 The block maxima method (BMM)

The basic idea in a picture based on losses $X_1, \dots, X_{12} \overset{\text{ind.}}{\sim} F \in \mathrm{MDA}(H_\xi)$:



Consider the maximal loss from each block and fit $H_{\xi,\mu,\sigma}$ to them.

Fitting the GEV distribution

- Suppose $(x_i)_{i\in\mathbb{N}}$ are realizations of $(X_i)_{i\in\mathbb{N}} \overset{\text{ind.}}{\sim} F \in \mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$, where F is unknown. By Fisher–Tippett–Gnedenko Theorem,

$$\mathbb{P}(M_n \le x) = \mathbb{P}((M_n - d_n)/c_n \le (x - d_n)/c_n) \underset{n \text{ large}}{\approx} H_{\xi,\mu=d_n,\sigma=c_n}(x).$$

- For fitting $\boldsymbol{\theta} = (\xi, \mu, \sigma)$, divide the realizations into m blocks of size n denoted by $M_{n1}, \dots, M_{nm}$ (e.g. daily log-returns $\Rightarrow$ monthly maxima)
- Assume the block size n to be sufficiently large so that (regardless of whether the underlying data are dependent or not), the block maxima can be considered independent.
- The density h_ξ of H_ξ is

$$h_\xi(x) = \begin{cases} (1+\xi x)^{-1/\xi - 1} H_\xi(x) I_{\{1+\xi x > 0\}}, & \text{if } \xi \neq 0, \\ e^{-x} H_0(x), & \text{if } \xi = 0. \end{cases}$$

The log-likelihood is thus

$$\ell(\boldsymbol{\theta}; M_{n1}, \ldots, M_{nm}) = \sum_{i=1}^{m} \log\Big(\frac{1}{\sigma} h_\xi\Big(\frac{M_{ni} - \mu}{\sigma}\Big)\Big).$$

Maximize w.r.t. $\boldsymbol{\theta} = (\xi, \mu, \sigma)$ to get $\hat{\boldsymbol{\theta}} = (\hat{\xi}, \hat{\mu}, \hat{\sigma})$.

Remark 5.8

1) Sufficiently many/large blocks require large amounts of data.
2) Bias and variance must be traded off (*bias-variance tradeoff*):
   - Block size $n \uparrow \Rightarrow$ GEV approximation more accurate $\Rightarrow$ bias $\downarrow$
   - Number of blocks $m \uparrow \Rightarrow$ more data for MLE $\Rightarrow$ variance $\downarrow$
3) There is no general best strategy for finding the optimal block size.

Return levels and return periods

(Approximately) $M_n \sim H_{\hat{\xi},\hat{\mu},\hat{\sigma}}$, so $\bar{H}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(r) = \mathbb{P}(M_n > r) = 1/k$ can be used to estimate the...

1) ... *k n-block return level* $r_{n,k}$, that is, the (smallest) r which is expected to be exceeded (at most) in one out of every k blocks of size n.
   - e.g., 10 year return level $r_{260,10}$ = (smallest) level exceeded in (at most) one out of every 10 years (where 260d $\approx$ 1y)
   - $r_{n,k} = H^{\leftarrow}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(1-1/k) \underset{\hat{\xi}\neq 0}{=} \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}((-\log(1-1/k))^{-\hat{\xi}} - 1)$
2) ... *return period* $k_{n,u}$ of the event $\{M_n > u\}$, that is, the smallest number of n-blocks for which we expect to see at least one n-block exceeding u.
   - $k_{n,u} = 1/\bar{H}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(u) \underset{\hat{\xi}\neq 0}{=} 1/(1-\exp(-(1+\hat{\xi}(x-\hat{\mu})/\hat{\sigma})^{-1/\hat{\xi}}))$
   - $k_{n,u}$ satisfies $r_{n,k_{n,u}} = u$

Example 5.9 (Block maxima analysis of S&P 500)

Suppose it is Friday 1987-10-16; the Friday before Black Monday (1987-10-19). The S&P 500 index fell by 9.12% this week. On that Friday alone the index is down 5.16%. We fit a GEV distribution to (bi)annual maxima of daily negative log-returns $X_t = -\log(S_t/S_{t-1})$ since 1960-01-01.

Analysis 1: Annual maxima ($m = 28$; including the latest from the incomplete year 1987): $\hat{\boldsymbol{\theta}} = (0.30, 0.02, 0.007) \Rightarrow$ Heavy-tailed Fréchet distribution (infinite fourth moment). The corresponding standard errors are $(0.22, 0.002, 0.001) \Rightarrow$ High uncertainty (m small) for estimating ξ.

Analysis 2: Biannual maxima ($m = 56$): $\hat{\boldsymbol{\theta}} = (0.34, 0.02, 0.006)$ with standard errors $(0.15, 0.0008, 0.0005) \Rightarrow$ Even heavier tails. In what follows we work with the annual maxima.

1) What is the probability that next year's maximal risk-factor change exceeds all previous ones? $1 - H_{\hat{\xi},\hat{\mu},\hat{\sigma}}(\text{“previous maxima”})$

2) Was a risk-factor change as on Black Monday foreseeable?
   - Based on data up to and including Friday 1987-10-16, the 10-year return level $r_{260,10}$ is estimated as $\hat{r}_{260,10} = 4.42\%$.
   - Index drop Black Monday: $20.47\% \Rightarrow X_{t+1} = 22.9\% \gg \hat{r}_{260,10}$.
   - One can show that 22.9% is in the 95% confidence interval of $r_{260,50}$ (estimated as $\hat{r}_{260,50} = 7.49\%$), but the 28 maxima are too few to get a reliable estimate of a once-in-50-years event.

3) Based on the available data, what is the (estimated) return period of a loss at least as large as on Black Monday?
   - The estimated return period $k_{260,0.229}$ is $\hat{k}_{260,0.229} = 1876$ years.
   - One can show that the 95% confidence interval encompasses everything from 45y to essentially never! $\Rightarrow$ Very high uncertainty!

$\Rightarrow$ On 1987-10-16 we did not have enough data to say anything meaningful about such an event. Quantifying such events is difficult.

5.2 Threshold exceedances

The BMM is wasteful of data (only the maxima of large blocks are used). It has been largely superseded in practice by methods based on threshold exceedances (*peaks-over-threshold (POT)* approach), where all data above a designated high threshold u are used.

5.2.1 Generalized Pareto distribution

Definition 5.10 (Generalized Pareto distribution (GPD))
The *generalized Pareto distribution (GPD)* is given by

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp(-x/\beta), & \text{if } \xi = 0, \end{cases}$$

where $\beta > 0$, and the support is $x \geq 0$ when $\xi \geq 0$ and $x \in [0, -\beta/\xi]$ when $\xi < 0$.

- ξ is known as *shape*; β as *scale*. Special cases:

  1) $\xi > 0$: $\mathrm{Par}(1/\xi, \beta/\xi)$

  2) $\xi = 0$: $\mathrm{Exp}(1/\beta)$

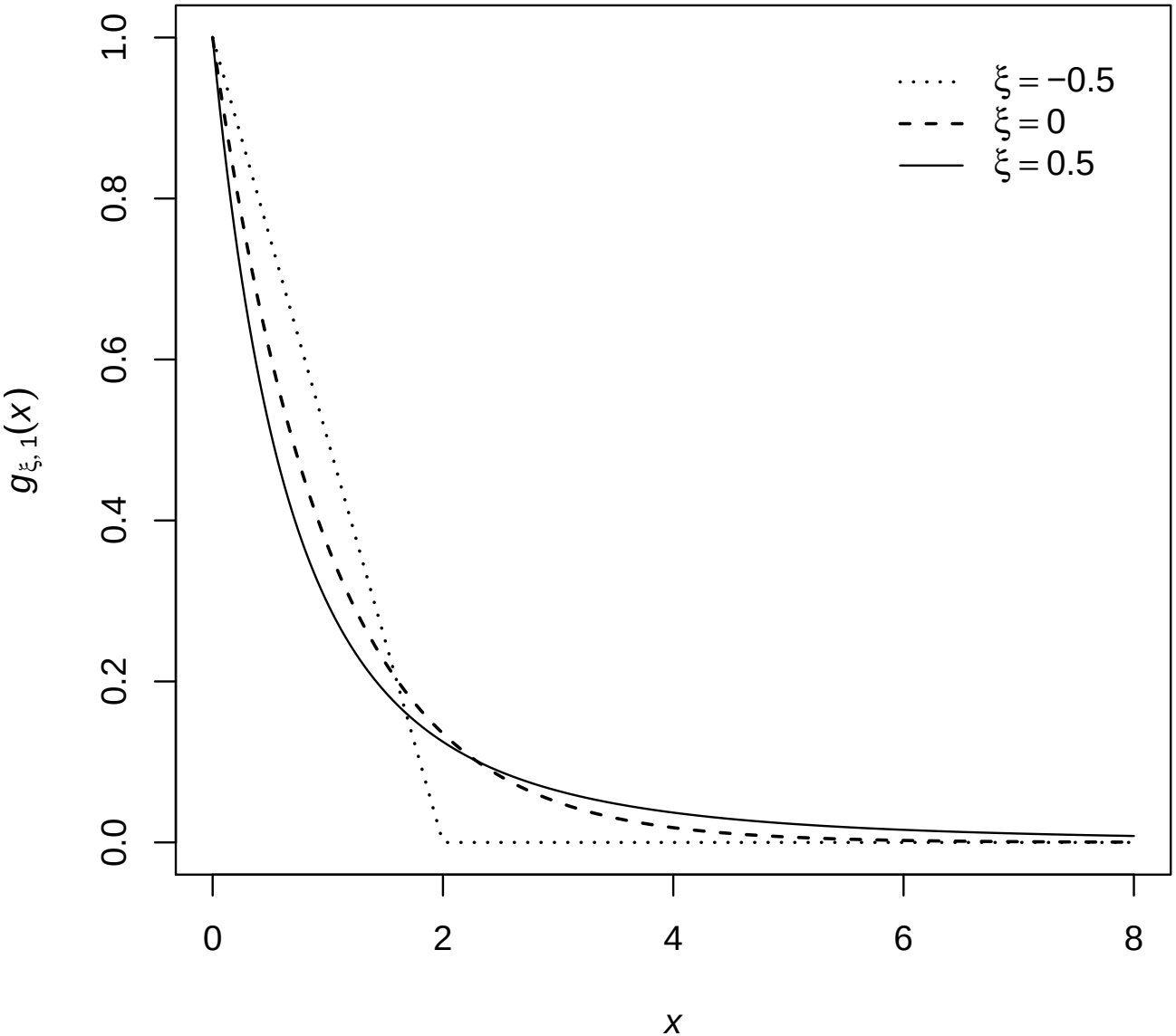
  3) $\xi < 0$: short-tailed Pareto type II distribution

- The larger ξ, the heavier tailed $G_{\xi,\beta}$ (if $\xi > 0$, $\mathbb{E}(X^k) = \infty$ iff $k \geq \frac{1}{\xi}$; if $\xi < 1$, then $\mathbb{E}X = \beta/(1-\xi)$).
- $G_{\xi,\beta} \in \mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$, (same ξ)
- The density $g_{\xi,\beta}$ of $G_{\xi,\beta}$ is given by

$$g_{\xi,\beta}(x) = \begin{cases} \frac{1}{\beta}(1+\xi x/\beta)^{-1/\xi-1}, & \text{if } \xi \neq 0, \\ \frac{1}{\beta}\exp(-x/\beta), & \text{if } \xi = 0, \end{cases}$$

  where $x \geq 0$ when $\xi \geq 0$ and $x \in [0, -\beta/\xi)$ when $\xi < 0$ (MLE!).

Density $g_{\xi,1}$ for $\xi \in \{-0.5, 0, 0.5\}$ (dotted, dashed, solid)



Definition 5.11 (Excess distribution over u, mean excess function)
Let $X \sim F$. The *excess distribution over the threshold u* is defined by

$$F_u(x) = \mathbb{P}(X - u \leq x \,|\, X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}, \quad x \in [0, x_F - u).$$

If $\mathbb{E}|X| < \infty$, the *mean excess function* is defined by

$$e(u) = \mathbb{E}(X - u \,|\, X > u) \quad \text{(i.e. the mean w.r.t. } F_u\text{)}$$

- **Interpretation:** F_u is the distribution of the excess loss $X - u$ over u, given that $X > u$. $e(u)$ is the mean of F_u as a function of u.
- One can show the useful formulas $e(u) = \int_0^{x_F - u} \bar{F}_u(x)\,\mathrm{d}x = \frac{\int_u^{x_F} \bar{F}(x)\,\mathrm{d}x}{\bar{F}(u)}$.

Example 5.12 (F_u, $e(u)$ for $\mathrm{Exp}(\lambda)$, $G_{\xi,\beta}$)

1) If F is $\mathrm{Exp}(\lambda)$, then $F_u(x) = 1 - e^{-\lambda x}$ and $e(u) = 1/\lambda = \mathbb{E}X$.
2) If F is $G_{\xi,\beta}$, then $F_u(x) = G_{\xi,\beta+\xi u}(x)$ and $e(u) = \frac{\beta+\xi u}{1-\xi}$ for all u : $\beta + \xi u > 0$, which is linear in u.

Theorem 5.13 (Pickands–Balkema–de Haan (1974/75))
There exists a positive, measurable function $\beta(u)$, such that

$$\lim_{u \uparrow x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0.$$

if and only if $F \in \mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$.

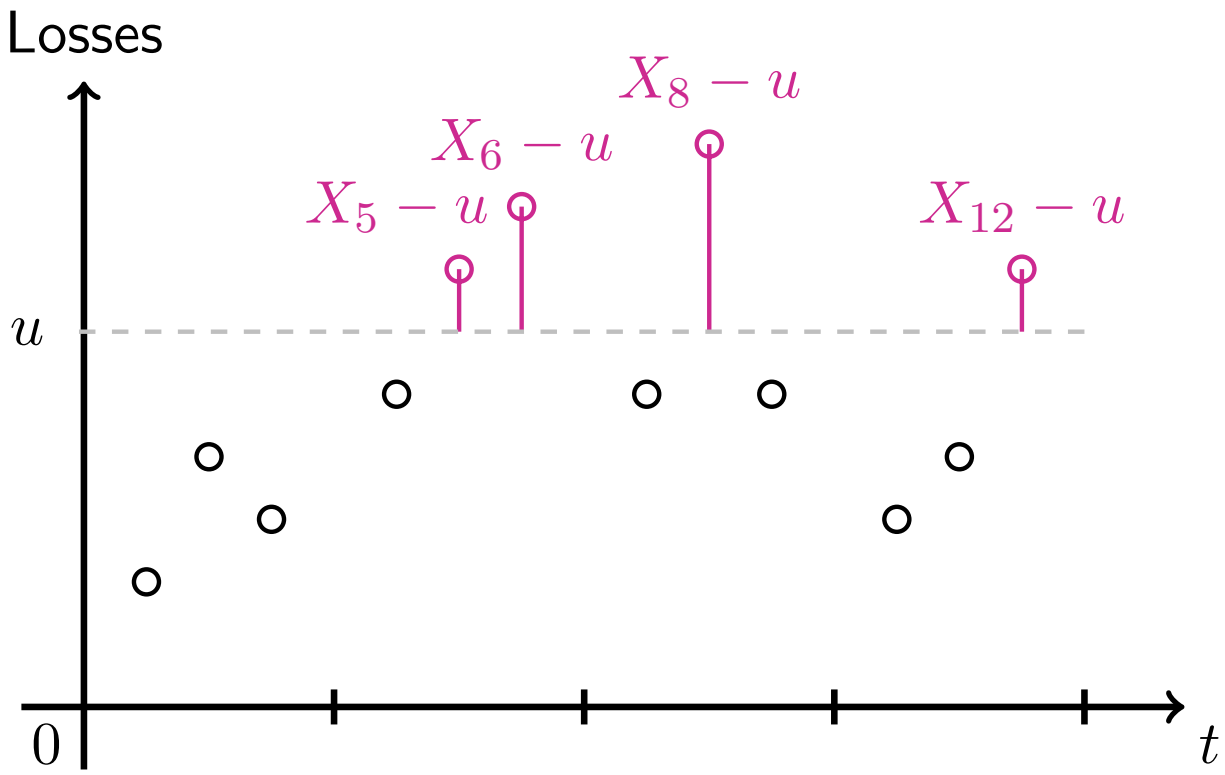
Proof. Non-trivial; see, e.g. Pickands (1975) and Balkema and de Haan (1974). □

Interpretation

- The GPD is the canonical df for excess losses over high u. This leads to the peaks-over-threshold method for modeling excess losses.
- The result is also a characterization of $\mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$. All $F \in \mathrm{MDA}(H_\xi)$ form a set of df for which the excess distribution converges to the GPD $G_{\xi,\beta}$ with the same ξ as in H_ξ when u is raised.

5.2.2 Modelling excess losses

The basic idea in a picture based on losses $X_1, \ldots, X_{12}$.



Consider all excesses over u and fit $G_{\xi,\beta}$ to them.

The peaks-over-threshold (POT) method

- Given losses $X_1, \dots, X_n \sim F \in \mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$, let

  1) $N_u = |\{i \in \{1, \dots, n\} : X_i > u\}|$ denote the *number of exceedances* over the (given; see later) threshold u;

  2) $\tilde{X}_1, \dots, \tilde{X}_{N_u}$ denote the *exceedances*; and

  3) $Y_k = \tilde{X}_k - u$, $k \in \{1, \dots, N_u\}$, the corresponding *excesses*.

- If $Y_1, \dots, Y_{N_u}$ are independent and (roughly) distributed as $G_{\xi,\beta}$, the log-likelihood is given by

$$\ell(\xi, \beta; Y_1, \dots, Y_{N_u}) = \sum_{k=1}^{N_u} \log g_{\xi,\beta}(Y_k)$$

$$= -N_u \log(\beta) - (1 + 1/\xi) \sum_{k=1}^{N_u} \log(1 + \xi Y_k/\beta)$$

  $\Rightarrow$ Maximize w.r.t. $\beta > 0$ and $1 + \xi Y_k/\beta > 0$ for all $k \in \{1, \dots, N_u\}$.

Excesses over higher thresholds

Once a model is fitted to F_u, we can infer a model for F_v, $v \geq u$.

Lemma 5.14
Assume, for some u, $F_u(x) = G_{\xi,\beta}(x)$ for $0 \leq x < x_F - u$. Then $F_v(x) = G_{\xi,\, \beta+\xi(v-u)}(x)$ for all $v \geq u$.

$\Rightarrow$ The excess distribution over $v \geq u$ remains GPD with the same ξ (and β growing linearly in v); makes sense for a limiting distribution for $u \uparrow$.

If it exists (so if $\xi < 1$), the mean excess function over v is given by

$$e(v) = \mathbb{E}(G_{\xi,\, \beta+\xi(v-u)}) = \frac{\beta + \xi(v-u)}{1-\xi} = \frac{\xi}{1-\xi} v + \frac{\beta - \xi u}{1-\xi}, \quad v \in [u, x_F), \tag{8}$$

where $x_F = \infty$ if $\xi \in [0, 1)$ and $x_F = u - \beta/\xi$ if $\xi < 0$. This forms the basis for a graphical method for choosing u.

Sample mean excess plot and choice of the threshold

Definition 5.15 (Sample mean excess function, mean excess plot) For $X_1, \ldots, X_n > 0$, the *sample mean excess function* is defined by

$$e_n(v) = \frac{\sum_{i=1}^n (X_i - v) I_{\{X_i > v\}}}{\sum_{i=1}^n I_{\{X_i > v\}}}, \quad v < X_{(n)}.$$

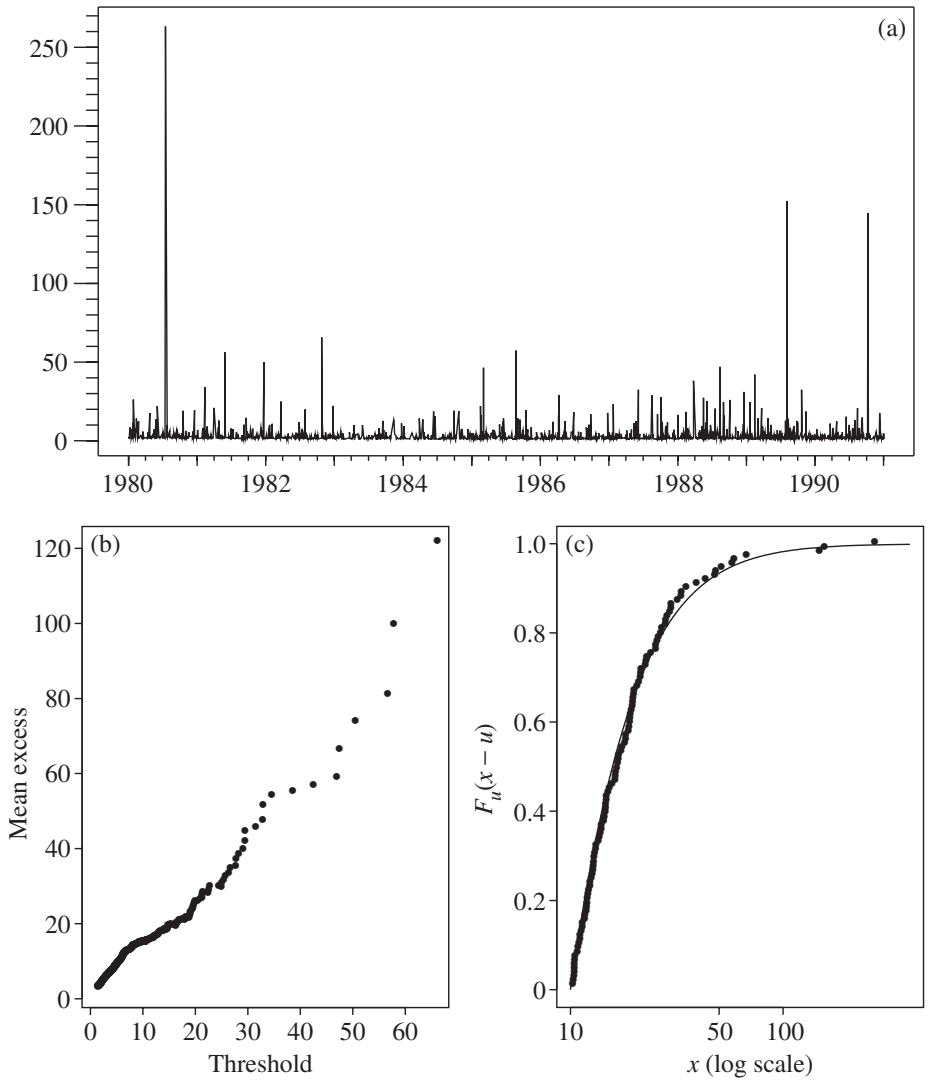
The *mean excess plot* is the plot of $\{(X_{(i)}, e_n(X_{(i)})) : 1 \leq i \leq n-1\}$, where $X_{(i)}$ denotes the ith order statistic.

- If the data supports the GPD model over u, $e_n(v)$ should become increasingly "linear" for higher values of $v \geq u$. Select u as the smallest point where $e_n(v)$, $v \geq u$, becomes linear; rule of thumb: 0.9-quantile.
- $e_n(v)$ is rarely perfectly linear. The choice of a good threshold u is as difficult as finding an adequate block size for the Block Maxima method. One should always analyze the data for several u.

Example 5.16 (Danish fire loss data)

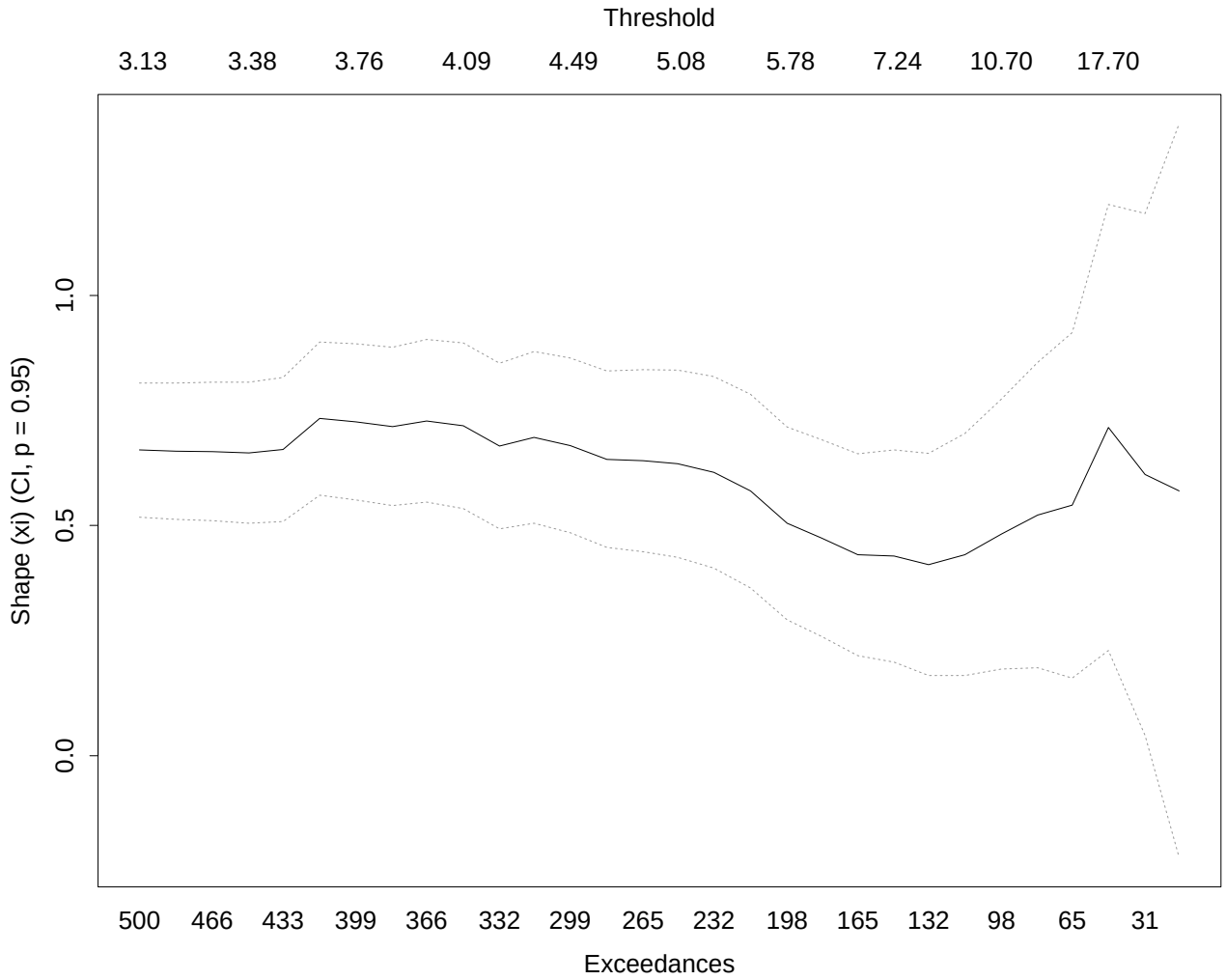
- 2156 fire insurance losses over 1M Danish kroner from 1980-01-03 to 1990-12-31; combined loss for a building and its contents, in some cases also a loss of business earnings. The losses are inflation adjusted to reflect values as of 1985.
- The sample mean excess function shows a "kink" below 10; "straightening out" above 10 $\Rightarrow$ Our choice is $u = 10$ (so 10M Danish kroner).
- MLE $(\hat{\xi}, \hat{\beta}) = (0.50, 7.0)$ (with standard errors $(0.14, 1.1)$)
  $\Rightarrow$ very heavy-tailed, infinite-variance model
- We can then estimate the expected loss given exceedance of 10M kroner or any higher threshold (via $e(v)$ in (8) based on $\hat{\xi}, \hat{\beta}$ and the chosen u), even beyond the data.
  $\Rightarrow$ EVT allows us to estimate "in the data" and then "scale up".

(a): Losses (> 1M; in M); (b): $e_n(u)$ ($\uparrow$); (c) $\hat{F}_{u,n}(x-u)$, $G_{\hat{\xi},\hat{\beta}}(x-u)$



$\Rightarrow$ Choose the threshold $u = 10$.

Sensitivity of the estimated shape parameter $\hat{\xi}$ to changes in u:



$\Rightarrow$ The higher u, the wider the confidence intervals (also support $u = 10$).

5.2.3 Modelling tails and measures of tail risk

- How can the fitted GPD model be used to estimate the tail of the loss distribution F and associated risk measures?
- Assume $F_u(x) = G_{\xi,\beta}(x)$ for $0 \le x < x_F - u$, $\xi \neq 0$ and some u.
- We obtain the following GPD-based formula for tail probabilities:

$$\begin{aligned}\bar{F}(x) &= \mathbb{P}(X > x) = \mathbb{P}(X > u)\mathbb{P}(X > x \,|\, X > u) \\ &= \bar{F}(u)\mathbb{P}(X - u > x - u \,|\, X > u) = \bar{F}(u)\bar{F}_u(x-u) \\ &= \bar{F}(u)\Big(1 + \xi\frac{x-u}{\beta}\Big)^{-1/\xi}, \quad x \ge u. \end{aligned} \tag{9}$$

- Assuming we know $\bar{F}(u)$, inverting this formula for $\alpha \ge F(u)$ leads to

$$\mathrm{VaR}_\alpha = F^{\leftarrow}(\alpha) = u + \frac{\beta}{\xi}\Big(\Big(\frac{1-\alpha}{\bar{F}(u)}\Big)^{-\xi} - 1\Big), \tag{10}$$

$$\mathrm{ES}_\alpha = \frac{\mathrm{VaR}_\alpha}{1-\xi} + \frac{\beta - \xi u}{1-\xi}, \quad \xi < 1. \tag{11}$$

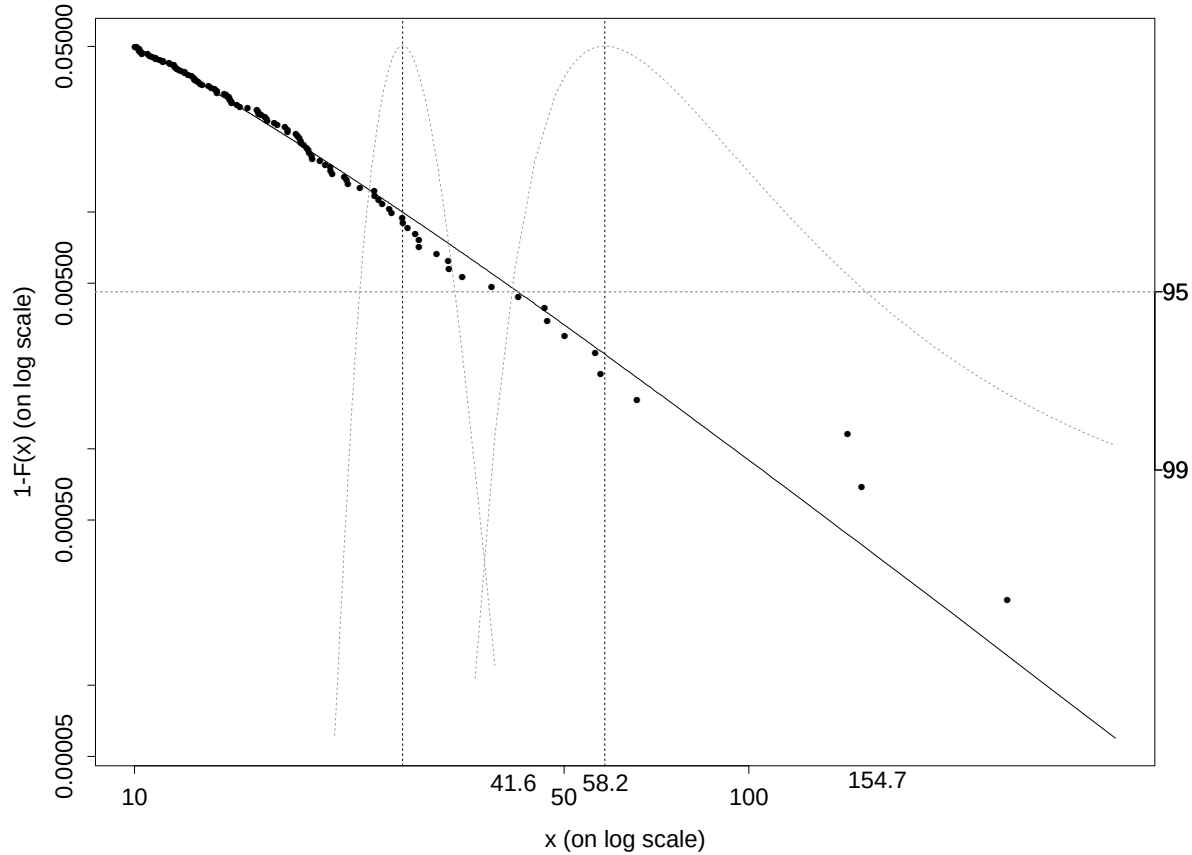
- $\bar{F}(x)$, VaR_α and ES_α are all of the form $g(\xi, \beta, \bar{F}(u))$. If we have sufficient samples above u, we obtain semi-parametric plug-in estimators via $g(\hat{\xi}, \hat{\beta}, N_u/n)$. We hope to gain over empirical estimators by using a kind of extrapolation based on the GPD for more extreme tail probabilities and risk measures.
- In this spirit, Smith (1987) proposed the *tail estimator*

$$\hat{\bar{F}}(x) = \frac{N_u}{n}\left(1 + \hat{\xi}\,\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}, \quad x \geq u \quad \text{(see (9))};$$

  also known as the *Smith estimator* (note that it is only valid for $x \geq u$). It faces a bias-variance tradeoff: If u is increased, the bias of parametrically estimating $\bar{F}_u(x-u)$ decreases, but the variance of it and the nonparametrically estimated $\bar{F}(u)$ increases.
- Similarly, semi-parametric GPD-based $\widehat{\mathrm{VaR}}_\alpha$, $\widehat{\mathrm{ES}}_\alpha$ for $\alpha \geq 1 - N_u/n$ can be obtained from (10), (11).

Example 5.17 (Danish fire loss data (continued))

Here are $\hat{\bar{F}}(x)$, $x \geq u$, $\widehat{\mathrm{VaR}}_{0.99}$, $\widehat{\mathrm{ES}}_{0.99}$ including confidence intervals.



Log-log scale often helpful: If $\bar{F}(x) = x^{-\alpha} L(x)$, $\log \bar{F}(x) = -\alpha \log(x) + \log L(x)$ which is approximately linear in $\log x$.

5.2.4 Conditional EVT for financial time series

- Assume $X_{t-n+1}, \ldots, X_t$ are negative log-returns generated by a strictly stationary time series process (X_t) of the form

$$X_t = \mu_t + \sigma_t Z_t,$$

  where μ_t and σ_t are $\mathcal{F}_{t-1}$-measurable and $Z_t \overset{\text{ind.}}{\sim} F_Z$; e.g. ARMA model with GARCH errors. Furthermore, let $Z \sim F_Z$.
- VaR_α^t and ES_α^t based on $F_{X_{t+1}|\mathcal{F}_t}$ are given by

$$\mathrm{VaR}_\alpha^t(X_{t+1}) = \mu_{t+1} + \sigma_{t+1}\,\mathrm{VaR}_\alpha(Z),$$
$$\mathrm{ES}_\alpha^t(X_{t+1}) = \mu_{t+1} + \sigma_{t+1}\,\mathrm{ES}_\alpha(Z).$$

- To obtain estimates $\widehat{\mathrm{VaR}}_\alpha^t(X_{t+1})$ and $\widehat{\mathrm{ES}}_\alpha^t(X_{t+1})$, proceed as follows:

  1) Fit an ARMA-GARCH model $\Rightarrow$ Estimates of μ_{t+1} and σ_{t+1}.
  2) Fit a GPD to the excesses corresponding to F_Z $\Rightarrow$ GPD-based estimates of $\mathrm{VaR}_\alpha(Z)$ (see (10)) and $\mathrm{ES}_\alpha(Z)$ (see (11)).

6 Multivariate models

6.1 Basics of multivariate modelling

6.2 Normal mixture distributions

6.3 Spherical and elliptical distributions

6.1 Basics of multivariate modelling

6.1.1 Random vectors and their distributions

Joint and marginal distributions

- Let $\boldsymbol{X} = (X_1, \dots, X_d) : \Omega \to \mathbb{R}^d$ be a d-dimensional *random vector* (representing risk-factor changes, risks, etc.).
- The *(joint) distribution function (df) F of $\boldsymbol{X}$* is

$$F(\boldsymbol{x}) = F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

- The *jth margin F_j of F* or *jth marginal df F_j of $\boldsymbol{X}$* is

$$\begin{aligned} F_j(x_j) &= \mathbb{P}(X_j \leq x_j) \\ &= \mathbb{P}(X_1 \leq \infty, \dots, X_{j-1} \leq \infty, X_j \leq x_j, X_{j+1} \leq \infty, \dots, X_d \leq \infty) \\ &= F(\infty, \dots, \infty, x_j, \infty, \dots, \infty), \quad x_j \in \mathbb{R},\ j \in \{1, \dots, d\}. \end{aligned}$$

(interpreted as a limit).

- Similarly for *k-dimensional margins*. Suppose we partition $\boldsymbol{X}$ into $(\boldsymbol{X}_1', \boldsymbol{X}_2')'$, where $\boldsymbol{X}_1 = (X_1, \dots, X_k)'$ and $\boldsymbol{X}_2 = (X_{k+1}, \dots, X_d)'$, then the marginal distribution function of $\boldsymbol{X}_1$ is

$$F_{\boldsymbol{X}_1}(\boldsymbol{x}_1) = \mathbb{P}(\boldsymbol{X}_1 \leq \boldsymbol{x}_1) = F(x_1, \dots, x_k, \infty, \dots, \infty).$$

- *F is absolutely continuous* if

$$F(\boldsymbol{x}) = \int_{-\infty}^{x_d} \cdots \int_{-\infty}^{x_1} f(z_1, \dots, z_d) \, \mathrm{d}z_1 \dots \mathrm{d}z_d = \int_{(-\infty, \boldsymbol{x}]} f(\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z}$$

for some $f \geq 0$ known as the *(joint) density of $\boldsymbol{X}$ (or F)*. Similarly, the *jth marginal df F_j is absolutely continuous* if $F_j(x) = \int_{-\infty}^{x} f_j(z) \, \mathrm{d}z$ for some $f_j \geq 0$ known as the *density of X_j (or F_j)*.

- Existence of a joint density $\Rightarrow$ Existence of marginal densities for all k-dimensional marginals, $1 \leq k \leq d-1$. The converse is false in general (counter-examples can be constructed with copulas; see Chapter 7).
- We sometimes work with the *survival function $\bar{F}$ of $\boldsymbol{X}$*,

$$\bar{F}(\boldsymbol{x}) = \bar{F}_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} > \boldsymbol{x}) = \mathbb{P}(X_1 > x_1, \dots, X_d > x_d), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

with corresponding *jth marginal survival function* $\bar{F}_j$

$$\bar{F}_j(x_j) = \mathbb{P}(X_j > x_j)$$
$$= \bar{F}(-\infty, \dots, -\infty, x_j, -\infty, \dots, -\infty), \quad x_j \in \mathbb{R},\ j \in \{1, \dots, d\}.$$

- Note that $\bar{F}(\boldsymbol{x}) \neq 1 - F(\boldsymbol{x})$ in general (unless $d = 1$), since, by the Law of Total Probability, $\bar{F}(x_1, x_2) = \mathbb{P}(X_1 > x_1, X_2 > x_2) = \mathbb{P}(X_1 > x_1) - \mathbb{P}(X_1 > x_1, X_2 \leq x_2) = 1 - \mathbb{P}(X_1 \leq x_1) - (\mathbb{P}(X_2 \leq x_2) - \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2)) = 1 - F_1(x_1) - F_2(x_2) + F(x_1, x_2) \neq 1 - F(x_1, x_2)$.

Independence

- A multivariate model for risks $\boldsymbol{X}$ in the form of a joint df, survival function or density, implicitly describes the *dependence* of $X_1, \dots, X_d$.
- $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ are *independent* if $F(\boldsymbol{x}_1, \boldsymbol{x}_2) = F_{\boldsymbol{X}_1}(\boldsymbol{x}_1) F_{\boldsymbol{X}_2}(\boldsymbol{x}_2)$ for all $\boldsymbol{x}_1, \boldsymbol{x}_2$ (if F has density f, then $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ are independent if $f(\boldsymbol{x}_1, \boldsymbol{x}_2) = f_{\boldsymbol{X}_1}(\boldsymbol{x}_1) f_{\boldsymbol{X}_2}(\boldsymbol{x}_2)$ for all $\boldsymbol{x}_1, \boldsymbol{x}_2$).

- The components $X_1, \dots, X_d$ of $\boldsymbol{X}$ are *(mutually) independent* if $F(\boldsymbol{x}) = \prod_{j=1}^d F_j(x_j)$ for all $\boldsymbol{x}$ (if F has density f, then $X_1, \dots, X_d$ are independent if $f(\boldsymbol{x}) = \prod_{j=1}^d f_j(x_j)$ for all $\boldsymbol{x}$).

Moments and characteristic function

- If $\mathbb{E}|X_j| < \infty$, $j \in \{1, \dots, d\}$, the *mean vector of* $\boldsymbol{X}$ is defined by

$$\mathbb{E}\boldsymbol{X} = (\mathbb{E}X_1, \dots, \mathbb{E}X_d).$$

  One can show: $X_1, \dots, X_d$ independent $\Rightarrow \mathbb{E}(X_1 \cdots X_d) = \prod_{j=1}^d \mathbb{E}(X_j)$
- If $\mathbb{E}(X_j^2) < \infty$ for all j, the *covariance matrix of* $\boldsymbol{X}$ is defined by

$$\mathrm{cov}(\boldsymbol{X}) = \mathbb{E}((\boldsymbol{X} - \mathbb{E}\boldsymbol{X})(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})').$$

  If we write $\Sigma = \mathrm{cov}(\boldsymbol{X})$, its (i,j)th element is

$$\begin{aligned}\sigma_{ij} = \Sigma_{ij} &= \mathrm{cov}(X_i, X_j) = \mathbb{E}((X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)) \\ &= \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j);\end{aligned}$$

  the diagonal elements are $\sigma_{jj} = \mathrm{var}(X_j)$, $j \in \{1, \dots, d\}$.

- X_1, X_2 independent $\overset{\Rightarrow}{\not\Leftarrow} \operatorname{cov}(X_1, X_2) = 0$ (counter-example: $X_1 \sim \mathrm{U}(-1,1)$, $X_2 = X_1^2 \Rightarrow \operatorname{cov}(X_1, X_2) = \mathbb{E}(X_1^3) - 0 \cdot \mathbb{E}(X_1^2) = 0$).
- If $\mathbb{E}(X_j^2) < \infty$, $j \in \{1, \dots, d\}$, the *correlation matrix of* $\boldsymbol{X}$ is defined by the matrix $\operatorname{corr}(\boldsymbol{X})$ with (i,j)th element

$$\operatorname{corr}(X_i, X_j) = \frac{\operatorname{cov}(X_i, X_j)}{\sqrt{\operatorname{var}(X_i)\operatorname{var}(X_j)}}, \quad i, j \in \{1, \dots, d\},$$

  which is in $[-1,1]$ with $\operatorname{corr}(X_i, X_j) = \pm 1$ if and only if $X_j \overset{\text{a.s.}}{=} aX_i + b$ for some $a \neq 0$ and $b \in \mathbb{R}$.
- **Some properties of $\mathbb{E}(\cdot)$ and $\operatorname{cov}(\cdot,\cdot)$:**

  1) For all $A \in \mathbb{R}^{k \times d}$, $\boldsymbol{b} \in \mathbb{R}^k$:
     - $\mathbb{E}(A\boldsymbol{X} + \boldsymbol{b}) = A\mathbb{E}\boldsymbol{X} + \boldsymbol{b}$;
     - $\operatorname{cov}(A\boldsymbol{X} + \boldsymbol{b}) = A \operatorname{cov}(\boldsymbol{X}) A' = A\Sigma A'$; if $k = 1$ ($A = \boldsymbol{a}'$),

$$\boldsymbol{a}'\Sigma\boldsymbol{a} = \operatorname{cov}(\boldsymbol{a}'\boldsymbol{X}) = \operatorname{var}(\boldsymbol{a}'\boldsymbol{X}) \geq 0, \quad \boldsymbol{a} \in \mathbb{R}^d, \tag{12}$$

     i.e. covariance matrices are *positive semidefinite*.

2) If Σ is a *positive definite matrix* (i.e. $\boldsymbol{a}'\Sigma\boldsymbol{a} > 0$ for all $\boldsymbol{a} \in \mathbb{R}^d \backslash \{\boldsymbol{0}\}$), one can show that Σ is invertible.

3) A symmetric, positive (semi)definite Σ can be written as

$$\Sigma = AA' \quad \textit{Cholesky decomposition} \tag{13}$$

for a lower triangular matrix A with $A_{jj} > 0$ ($A_{jj} \geq 0$) for all j. A is known as *Cholesky factor* (and is also denoted by $\Sigma^{1/2}$).

- Properties of $\boldsymbol{X}$ can often be shown with the *characteristic function (cf)*

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}(\exp(i\boldsymbol{t}'\boldsymbol{X})), \quad \boldsymbol{t} \in \mathbb{R}^d.$$

$X_1, \ldots, X_d$ are independent $\Leftrightarrow \phi_{\boldsymbol{X}}(\boldsymbol{t}) = \prod_{j=1}^d \phi_{X_j}(t_j)$ for all $\boldsymbol{t}$.

Proposition 6.1 (Characterization of covariance matrices)
A symmetric matrix Σ is a covariance matrix if and only if it is positive semidefinite.

6.1.2 Standard estimators of covariance and correlation

- Assume $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n \sim F$ (daily/weekly/monthly/yearly risk-factor changes) are serially uncorrelated (i.e. multivariate white noise) with $\boldsymbol{\mu} := \mathbb{E}\boldsymbol{X}_1$, $\Sigma := \operatorname{cov} \boldsymbol{X}_1$ and $P = \operatorname{corr}(\boldsymbol{X}_1)$.
- Standard estimators of $\boldsymbol{\mu}, \Sigma, P$ are

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \quad (\textit{sample mean})$$

$$S = \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})' \ (\textit{sample covariance matrix})$$

$$R = (R_{ij}) \text{ for } R_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}} \ (\textit{sample correlation matrix})$$

- Under joint normality (F multivariate normal), $\bar{\boldsymbol{X}}$, S and R are also MLEs. S is biased, but an unbiased version can be obtained by

$$S_n = \frac{n}{n-1} S.$$

6.1.3 The multivariate normal distribution

Definition 6.2 (Multivariate normal distribution)
$\boldsymbol{X} = (X_1, \dots, X_d)$ has a *multivariate normal* (or *Gaussian*) *distribution* if

$$\boldsymbol{X} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + A\boldsymbol{Z}, \tag{14}$$

where $\boldsymbol{Z} = (Z_1, \dots, Z_k)$, $Z_l \overset{\text{ind.}}{\sim} \mathrm{N}(0,1)$, $A \in \mathbb{R}^{d \times k}$, $\boldsymbol{\mu} \in \mathbb{R}^d$.

- Typically $k = d$
- $\mathbb{E}\boldsymbol{X} = \boldsymbol{\mu} + A\mathbb{E}\boldsymbol{Z} = \boldsymbol{\mu}$
- $\mathrm{cov}(\boldsymbol{X}) = \mathrm{cov}(\boldsymbol{\mu} + A\boldsymbol{Z}) = A\,\mathrm{cov}(\boldsymbol{Z})A' = AA' =: \Sigma$

Proposition 6.3 (Cf of the multivariate normal distribution)
Let $\boldsymbol{X}$ be as in (14) and $\Sigma = AA'$. Then the cf of $\boldsymbol{X}$ is

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}(\exp(i\boldsymbol{t}'\boldsymbol{X})) = \exp\left(i\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right), \quad \boldsymbol{t} \in \mathbb{R}^d.$$

- We see that the multivariate normal distribution is characterized by $\boldsymbol{\mu}$ and Σ, hence the notation $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$.
- Characterization ("$\Rightarrow$" via uniqueness of cfs):

$$\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma) \quad \Longleftrightarrow \quad \boldsymbol{a}'\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\Sigma\boldsymbol{a}) \quad \forall\, \boldsymbol{a} \in \mathbb{R}^d.$$

- Consequences:
  - $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma) \overset{\boldsymbol{a}=\boldsymbol{e}_j}{\underset{\not\Leftarrow}{\Rightarrow}} X_j \sim \mathrm{N}(\mu_j, \Sigma_{jj}), \quad j \in \{1, \ldots, d\}.$
  - $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma) \overset{\boldsymbol{a}=\mathbf{1}}{\Rightarrow} \sum_{j=1}^d X_j \sim \mathrm{N}(\sum_{j=1}^d \mu_j,\ \sum_{i,j} \Sigma_{ij})$.

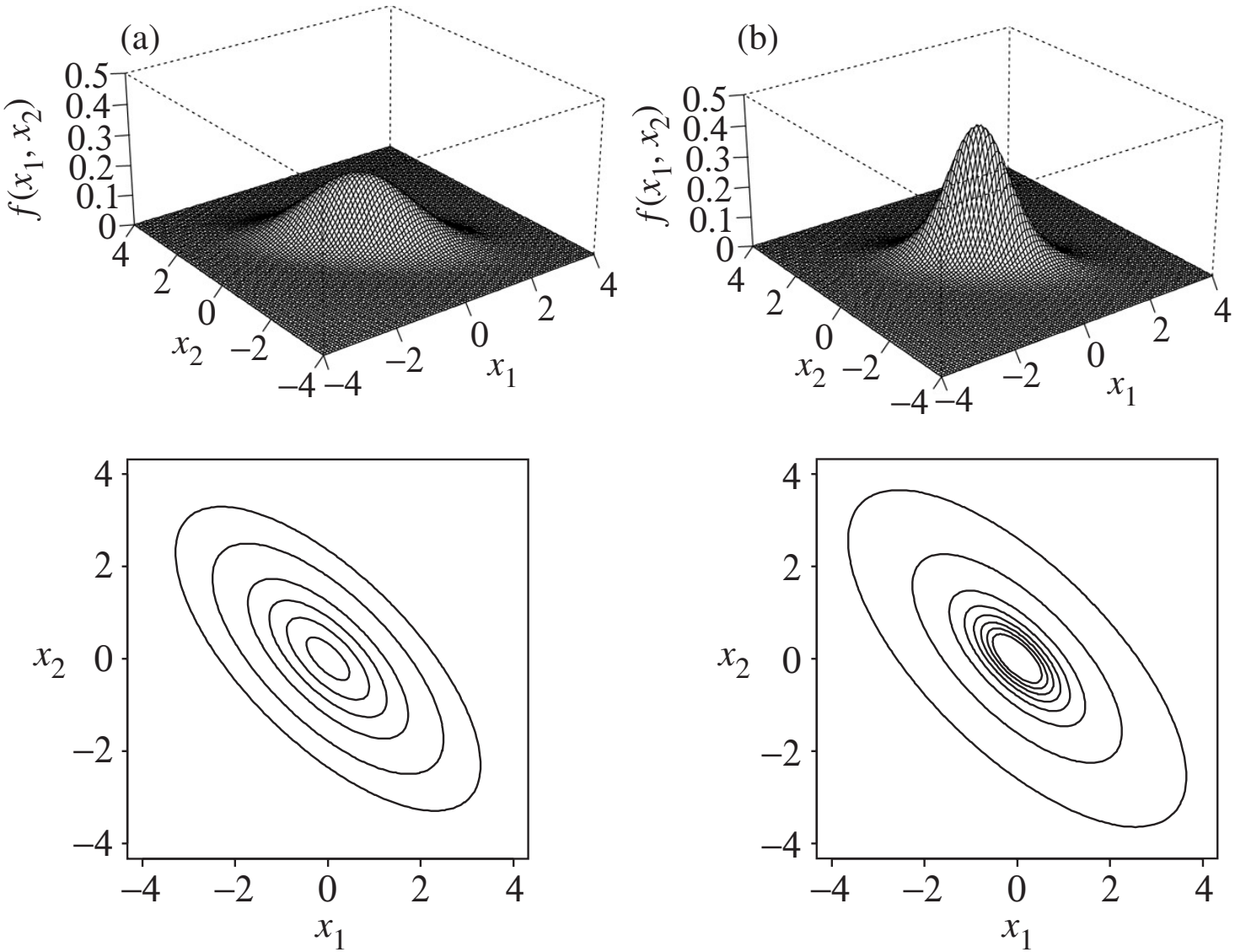
Proposition 6.4 (Density)

Let $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ with $\operatorname{rank} A = k = d$ ($\Rightarrow \Sigma$ pos. definite, invertible). By the density transformation theorem, $\boldsymbol{X}$ can be shown to have density

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

Consequences:

- Sets of the form $S_c = \{\boldsymbol{x} \in \mathbb{R}^d : (\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = c\}$, $c > 0$, describe points of equal density. Contours of equal density are thus ellipsoids.
- The components of $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ are mutually independent if and only if Σ is diagonal, i.e. if and only if the components of $\boldsymbol{X}$ are uncorrelated.



Left: $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ for $\boldsymbol{\mu} = \binom{0}{0}$, $\Sigma = \begin{pmatrix} 1 & -0.7 \\ -0.7 & 1 \end{pmatrix}$; Right: $t_\nu(\boldsymbol{\mu}, \frac{\nu-2}{\nu}\Sigma)$, $\nu = 4$, (same mean and covariance matrix as on the left-hand side)

The definition of $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ in terms of a stochastic representation ($\boldsymbol{X} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + A\boldsymbol{Z}$) directly justifies the following sampling algorithm.

Algorithm 6.5 (Sampling $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$)
Let $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ with Σ symmetric and positive definite.

1) Compute the Cholesky factor A of Σ; see, e.g. Press et al. (1992).
2) Generate $Z_j \overset{\text{ind.}}{\sim} \mathrm{N}(0,1)$, $j \in \{1, \dots, d\}$.
3) Return $\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Z}$, where $\boldsymbol{Z}a = (Z_1, \dots, Z_d)$.

Further useful properties of multivariate normal distributions

- **Linear combinations**
  If $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ and $B \in \mathbb{R}^{k \times d}, \boldsymbol{b} \in \mathbb{R}^k$, then

$$\begin{aligned} B\boldsymbol{X} + \boldsymbol{b} &= B(\boldsymbol{\mu} + A\boldsymbol{Z}) + \boldsymbol{b} = (B\boldsymbol{\mu} + \boldsymbol{b}) + BA\boldsymbol{Z} \\ &\sim \mathrm{N}_k(B\boldsymbol{\mu} + \boldsymbol{b},\ BA(BA)') = \mathrm{N}_k(B\boldsymbol{\mu} + \boldsymbol{b},\ B\Sigma B'). \end{aligned}$$

  Special case (see var.-cov. method): $\boldsymbol{b}'\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{b}'\boldsymbol{\mu}, \boldsymbol{b}'\Sigma\boldsymbol{b})$.

- **Marginal dfs**

  Let $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ and write $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')$, where $\boldsymbol{X}_1 \in \mathbb{R}^k$, $\boldsymbol{X}_2 \in \mathbb{R}^{d-k}$, and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Then

$$\boldsymbol{X}_1 \sim \mathrm{N}_k(\boldsymbol{\mu}_1, \Sigma_{11}) \quad \text{and} \quad \boldsymbol{X}_2 \sim \mathrm{N}_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22}).$$

  Proof. Choose $B = \left(\begin{smallmatrix} I_k & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 0 & 0 \\ 0 & I_{d-k} \end{smallmatrix}\right)$, respectively.

- **Quadratic forms**

  Let $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ and Σ be positive definite with Cholesky factor A. Furthermore, let $\boldsymbol{Z} = A^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$. Then $\boldsymbol{Z} \sim \mathrm{N}_d(\boldsymbol{0}, I_d)$. Moreover,

$$(\boldsymbol{X} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) = \boldsymbol{Z}' \boldsymbol{Z} \sim \chi_d^2, \tag{15}$$

  which is useful for (goodness-of-fit) testing of $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$: We can check whether the *squared Mahalanobis distances* $D_i^2 = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})' S^{-1} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})$, $i \in \{1, \ldots, n\}$, form a(n approximate) sample from χ_d^2.

6.1.4 Testing multivariate normality

- For testing univariate normality, all tests of Section 3.1.2 can be applied.
- Now consider multivariate normality.

$$\boldsymbol{X}_1, \dots, \boldsymbol{X}_n \overset{\text{ind.}}{\sim} \mathrm{N}_d(\boldsymbol{\mu}, \Sigma) \;\Rightarrow\; \boldsymbol{a}'\boldsymbol{X}_1, \dots, \boldsymbol{a}'\boldsymbol{X}_n \overset{\text{ind.}}{\sim} \mathrm{N}(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\Sigma\boldsymbol{a}).$$

  This can be tested statistically (for some $\boldsymbol{a}$) with various goodness-of-fit tests (e.g. Q-Q plots) used for univariate normality . Alternatively, (15) can be used to test joint normality (see Mardia's test below).

- Multivariate Shapiro–Wilk
- Mardia's test
  - According to (15), if $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ with Σ positive definite, then $(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2_d$ (can approx. be used in a Q-Q plot).
  - Let $D_i^2 = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})'S^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})$ denote the *squared Mahalanobis distances* and $D_{ij} = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})'S^{-1}(\boldsymbol{X}_j - \bar{\boldsymbol{X}})$ the *Mahalanobis angles*.

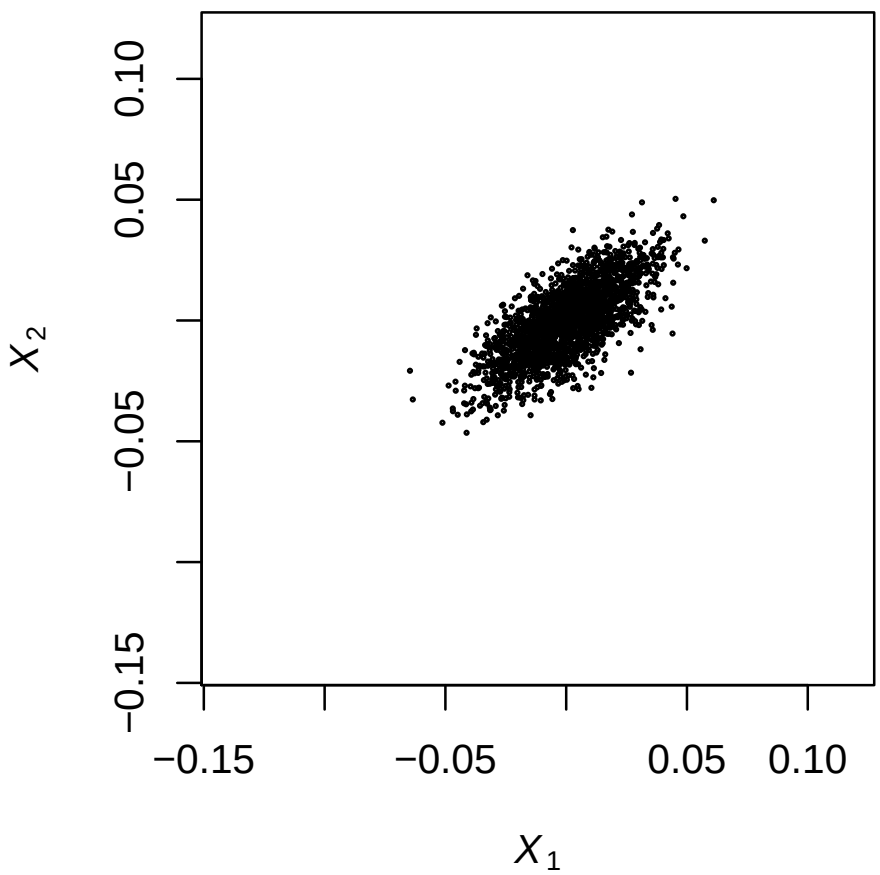
- Let $b_d = \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n D_{ij}^3$ and $k_d = \frac{1}{n}\sum_{i=1}^n D_i^4$. Under the null hypothesis one can show that asymptotically for $n \to \infty$,

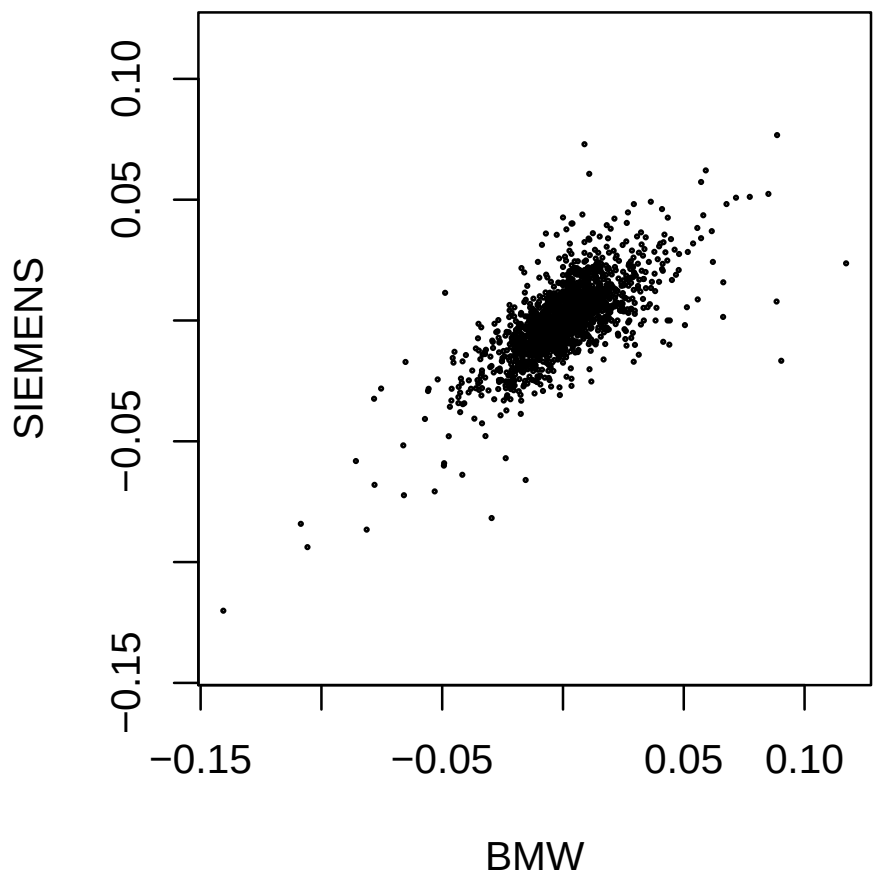
$$\frac{n}{6} b_d \sim \chi^2_{d(d+1)(d+2)/6}, \qquad \frac{k_d - d(d+2)}{\sqrt{8d(d+2)/n}} \sim \mathrm{N}(0,1),$$

  which can be used for testing; see Joenssen and Vogel (2014).

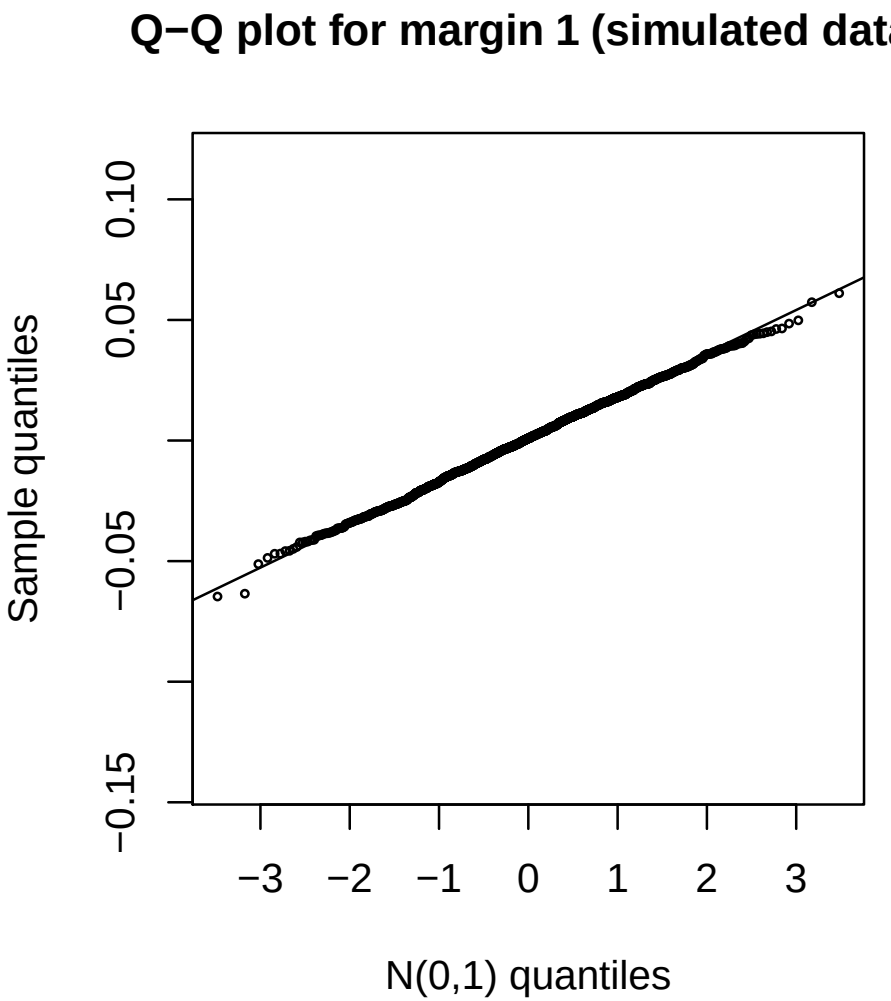
Example 6.6 (Simulated data vs BMW–Siemens)

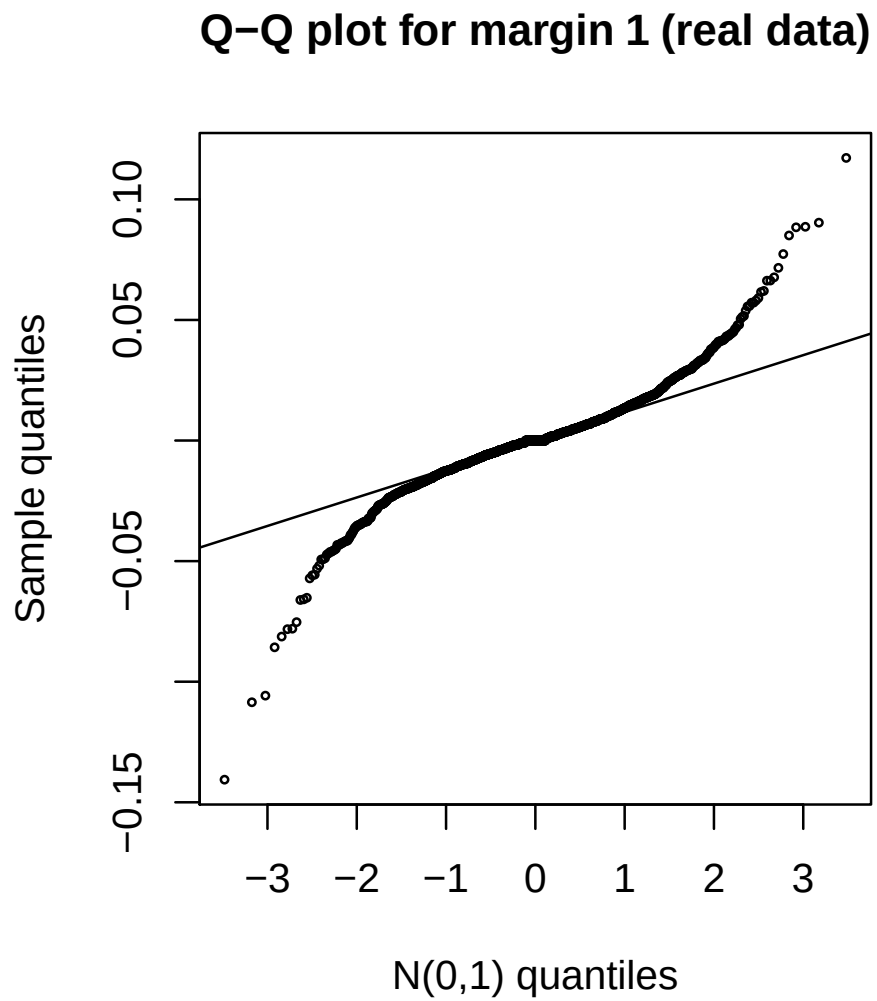
Is the BMW–Siemens data (see Section 3.2.2) jointly normal?



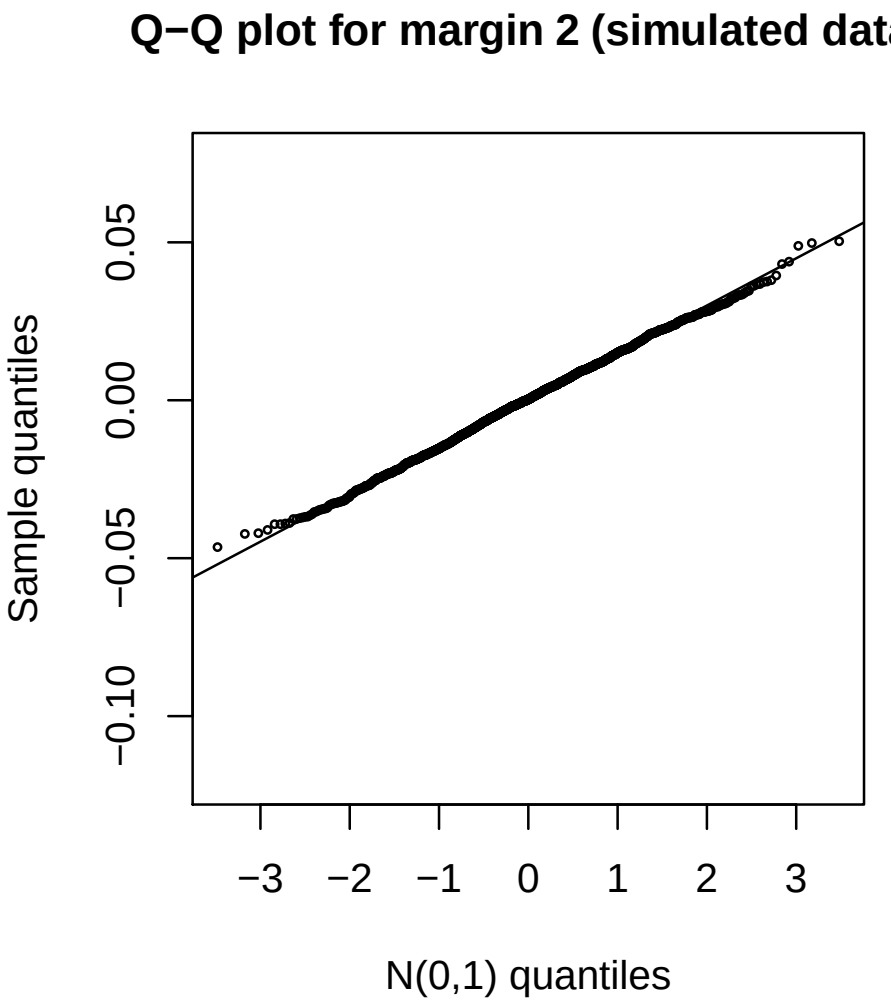


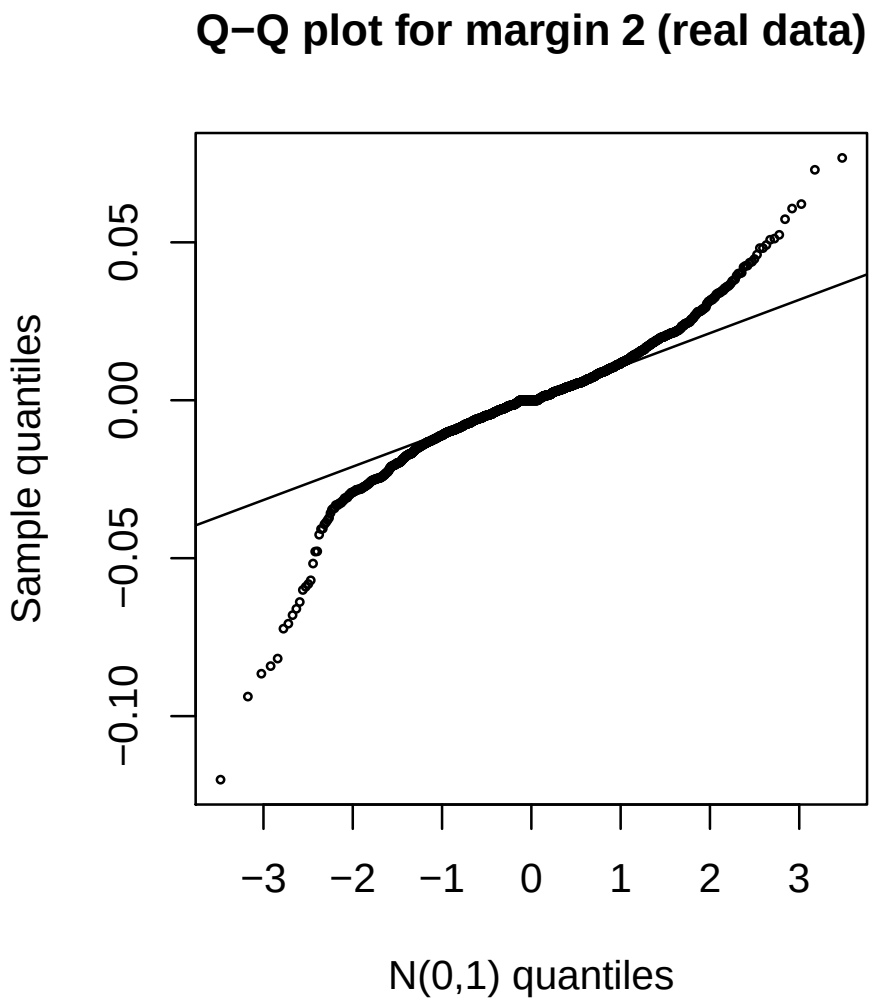
Considering the first margin only:



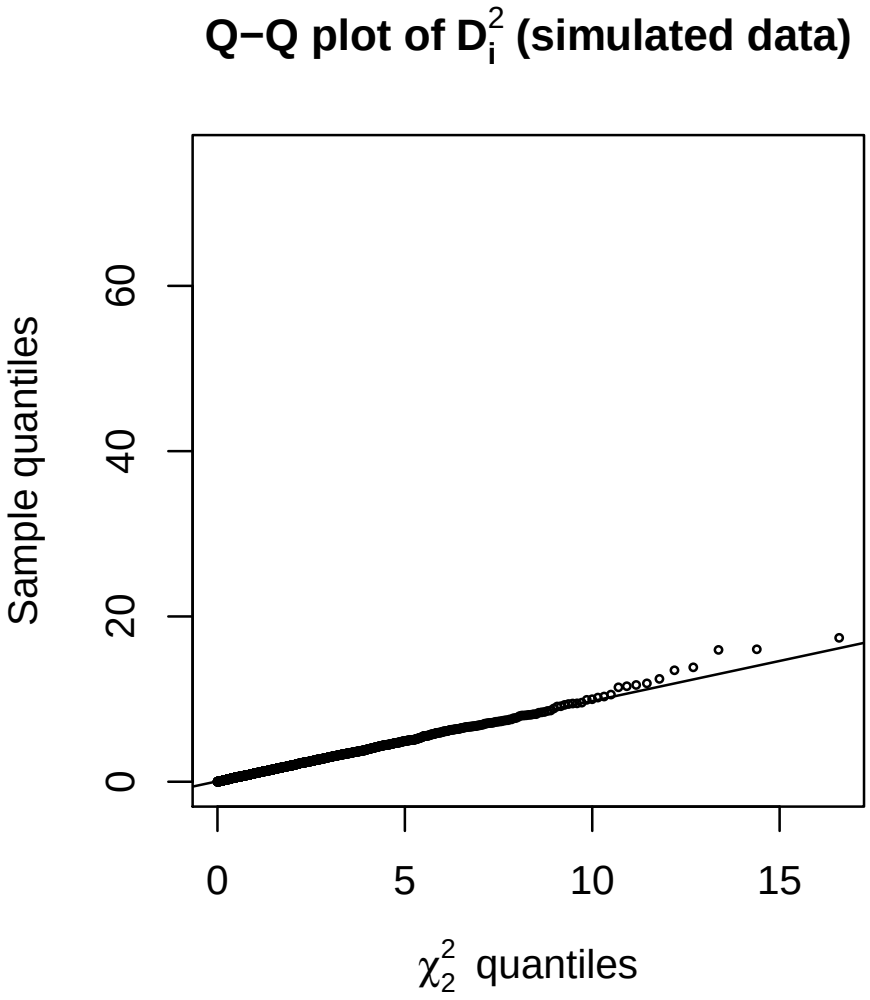


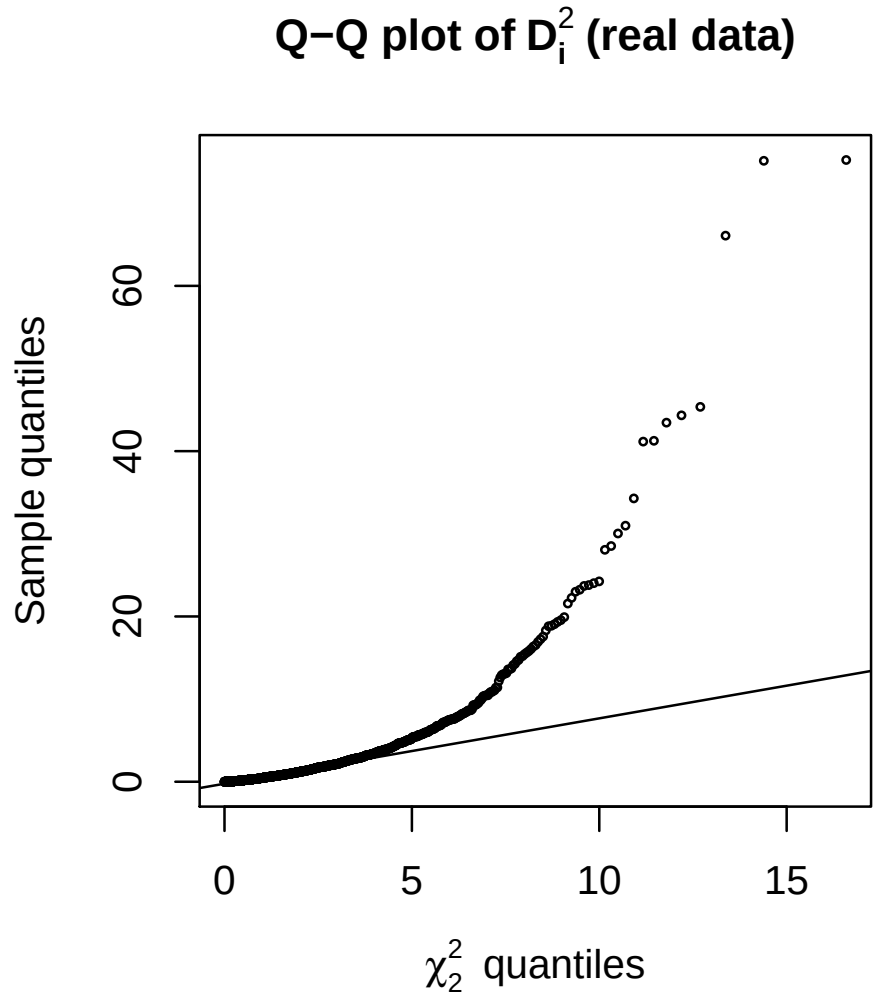
Considering the second margin only:





Q-Q plot of the simulated (left) or real (right) D_i^2's against a χ_2^2:





Advantages of $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$

- Distribution is determined by $\boldsymbol{\mu}$ and Σ.
- Inference is thus "easy".
- Linear combinations are normal ($\Rightarrow \mathrm{VaR}_\alpha$ and ES_α calculations for portfolios are easy).
- Marginal distributions are normal.
- Conditional distributions are normal.
- Quadratic forms are (theoretically) chi-squared.
- Convolutions are normal.
- Sampling is straightforward.
- Independence and uncorrelatedness are equivalent.

Drawbacks of $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ for modelling risk-factor changes

1) Tails of univariate (normal) margins are too thin (generate too few extreme events).
2) Joint tails are too thin (too few joint extreme events). $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ cannot capture the notion of tail dependence (see Chapters 3 and 7).
3) Strong symmetry known as radial symmetry: $\boldsymbol{X}$ is *radially symmetric about* $\boldsymbol{\mu}$ if $\boldsymbol{X} - \boldsymbol{\mu} \overset{\mathrm{d}}{=} \boldsymbol{\mu} - \boldsymbol{X}$. This is true for $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ since $\boldsymbol{Z} \overset{\mathrm{d}}{=} -\boldsymbol{Z}$.

Short outlook:

- Normal variance mixtures (or, more generally, elliptical distributions) can address 1) and 2) while sharing many of the desirable properties of $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$.
- Normal mean-variance mixtures can also address 3) (but at the expense of ellipticality and thus tractability in comparison to $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$).

6.2 Normal mixture distributions

Idea: Randomize Σ (and possibly $\boldsymbol{\mu}$) with a non-negative rv W.

6.2.1 Normal variance mixtures

Definition 6.7 (Multivariate normal variance mixtures)
The random vector $\boldsymbol{X}$ has a (multivariate) *normal variance mixture distribution* if

$$\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \sqrt{W} A \boldsymbol{Z},$$

where $\boldsymbol{Z} \sim \mathrm{N}_k(\boldsymbol{0}, I_k)$, $W \geq 0$ is a rv independent of $\boldsymbol{Z}$, $A \in \mathbb{R}^{d \times k}$, and $\boldsymbol{\mu} \in \mathbb{R}^d$. $\boldsymbol{\mu}$ is called *location vector* and $\Sigma = AA'$ *scale* (or *dispersion*) *matrix*.

Observe that $(\boldsymbol{X} \mid W = w) \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \sqrt{w} A \boldsymbol{Z} = \mathrm{N}_d(\boldsymbol{\mu}, wAA') = \mathrm{N}_d(\boldsymbol{\mu}, w\Sigma)$; or $(\boldsymbol{X} \mid W) \stackrel{\mathrm{d}}{=} \mathrm{N}_d(\boldsymbol{\mu}, W\Sigma)$. W can be interpreted as a shock affecting the variances of all risk factors.

Properties of multivariate normal variance mixtures

Let $\boldsymbol{X} = \boldsymbol{\mu} + \sqrt{W} A\boldsymbol{Z}$ and $\boldsymbol{Y} = \boldsymbol{\mu} + A\boldsymbol{Z}$. Assume that $\mathrm{rank}(A) = d \leq k$ and that Σ is positive definite.

- If $\mathbb{E}\sqrt{W} < \infty$, then $\mathbb{E}(\boldsymbol{X}) \overset{\text{ind.}}{=} \boldsymbol{\mu} + \mathbb{E}(\sqrt{W}) A \mathbb{E}(\boldsymbol{Z}) = \boldsymbol{\mu} + \boldsymbol{0} = \boldsymbol{\mu} = \mathbb{E}\boldsymbol{Y}$
- If $\mathbb{E}W < \infty$, then

$$\begin{aligned} \mathrm{cov}(\boldsymbol{X}) &= \mathrm{cov}(\sqrt{W} A\boldsymbol{Z}) = \mathbb{E}((\sqrt{W} A\boldsymbol{Z})(\sqrt{W} A\boldsymbol{Z})') \\ &\overset{\text{ind.}}{=} \mathbb{E}(W) \cdot \mathbb{E}(A\boldsymbol{Z}\boldsymbol{Z}'A') = \mathbb{E}(W) \cdot A\mathbb{E}(\boldsymbol{Z}\boldsymbol{Z}')A' \\ &= \mathbb{E}(W) A I_k A' = \mathbb{E}(W)\Sigma \underset{\text{in general}}{\neq} \Sigma \quad (= \mathrm{cov}(\boldsymbol{Y})) \end{aligned}$$

- However, if they exist (i.e. if $\mathbb{E}W < \infty$) $\mathrm{corr}(\boldsymbol{X}) = \mathrm{corr}(\boldsymbol{Y})$ since

$$\begin{aligned} \mathrm{corr}(X_i, X_j) &= \frac{\mathrm{cov}(X_i, X_j)}{\sqrt{\mathrm{var}(X_i)\,\mathrm{var}(X_j)}} = \frac{\mathbb{E}(W)\Sigma_{ij}}{\sqrt{\mathbb{E}(W)\Sigma_{ii}\mathbb{E}(W)\Sigma_{jj}}} \\ &= \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}} = \mathrm{corr}(Y_i, Y_j), \quad i, j \in \{1, \ldots, d\}. \end{aligned}$$

- **Characteristic function:** Recall: If $\boldsymbol{Y} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$, then $\phi_{\boldsymbol{Y}}(\boldsymbol{t}) = \exp(i\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t})$. The cf of a multivariate normal variance mixtures is

$$\begin{aligned}\phi_{\boldsymbol{X}}(\boldsymbol{t}) &= \mathbb{E}(\exp(i\boldsymbol{t}'\boldsymbol{X})) = \mathbb{E}(\,\mathbb{E}(\exp(i\boldsymbol{t}'\boldsymbol{X}) \,|\, W)\,) \\ &= \mathbb{E}(\exp(i\boldsymbol{t}'\boldsymbol{\mu} - \tfrac{1}{2}W\boldsymbol{t}'\Sigma\boldsymbol{t})) = \exp(i\boldsymbol{t}'\boldsymbol{\mu})\mathbb{E}(\exp(-W\tfrac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t})).\end{aligned}$$

  This depends on the *Laplace-Stieltjes transform* $\hat{F}_W(\theta) = \mathbb{E}(\exp(-\theta W)) = \int_0^\infty e^{-\theta w}\,\mathrm{d}F_W(w)$ of F_W. We thus introduce the notation $\boldsymbol{X} \sim M_d(\boldsymbol{\mu}, \Sigma, \hat{F}_W)$ for a d-dimensional multivariate normal variance mixture.

- **Density:** If Σ is positive definite, $\mathbb{P}(W = 0) = 0$, the density of $\boldsymbol{X}$ is

$$\begin{aligned}f_{\boldsymbol{X}}(\boldsymbol{x}) &= \int_0^\infty f_{\boldsymbol{X}|W}(\boldsymbol{x} \,|\, w)\,\mathrm{d}F_W(w) \\ &= \int_0^\infty \frac{1}{(2\pi)^{d/2}w^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2w}\right)\mathrm{d}F_W(w).\end{aligned}$$

  $\Rightarrow$ Only depends on $\boldsymbol{x}$ through $(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$.

  $\Rightarrow$ Multivariate normal variance mixtures are elliptical distributions.

If Σ is diagonal and $\mathbb{E}W < \infty$, $\boldsymbol{X}$ is uncorrelated (as $\mathrm{cov}(\boldsymbol{X}) = \mathbb{E}(W)\Sigma$) but not independent unless W is constant a.s. (see stoch. representation).

- **Linear combinations:** For $\boldsymbol{X} \sim M_d(\boldsymbol{\mu}, \Sigma, \hat{F}_W)$ and $\boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{b}$, where $B \in \mathbb{R}^{k \times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$, we have $\boldsymbol{Y} \sim M_k(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B', \hat{F}_W)$; this can be shown via cfs.

- **Sampling:**

Algorithm 6.8 (Simulation of $\boldsymbol{X} = \boldsymbol{\mu} + \sqrt{W}A\boldsymbol{Z} \sim M_d(\boldsymbol{\mu}, \Sigma, \hat{F}_W)$)

1) Generate $\boldsymbol{Z} \sim \mathrm{N}_d(\boldsymbol{0}, I_d)$.
2) Generate $W \sim F_W$ (with LS transform $\hat{F}_W$), independent of $\boldsymbol{Z}$.
3) Compute the Cholesky factor A (such that $AA' = \Sigma$).
4) Return $\boldsymbol{X} = \boldsymbol{\mu} + \sqrt{W}A\boldsymbol{Z}$.

Examples of multivariate normal variance mixtures

- **Multivariate normal distribution**
  $W = 1$ a.s. (degenerate case)
- **Two point mixture**

$$W = \begin{cases} w_1 \text{ with probability } p, \\ w_2 \text{ with probability } 1-p \end{cases} \qquad w_1,\ w_2 > 0,\ w_1 \neq w_2.$$

  Can be used to model ordinary and stress regimes; extends to k regimes.
- **Symmetric generalised hyperbolic distribution**
  W has a generalised inverse Gaussian distribution (GIG); see MFE (2015, p. 187).
- **Multivariate t distribution**
  W has an inverse gamma distribution $W = 1/V$ for $V \sim \Gamma(\nu/2, \nu/2)$.
  - $\mathbb{E}(W) = \frac{\nu}{\nu-2} \Rightarrow \text{cov}(\boldsymbol{X}) = \frac{\nu}{\nu-2}\Sigma$. For finite variances/correlations, $\nu > 2$ is required. For finite mean, $\nu > 1$ is required.

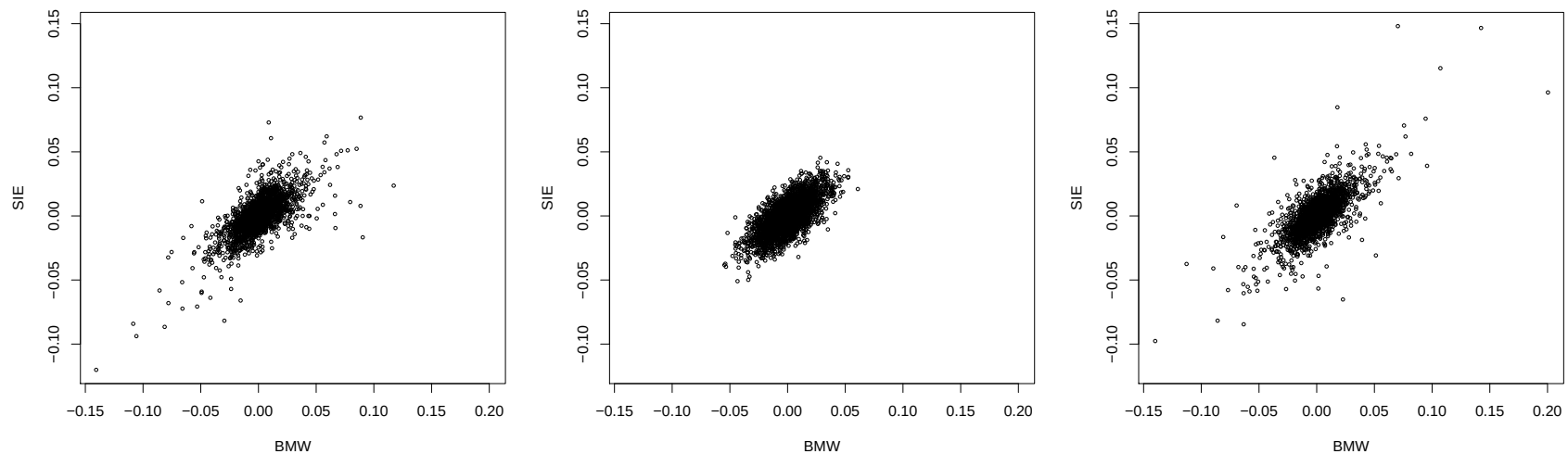
- The density of the multivariate t distribution is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)(\nu\pi)^{d/2}|\Sigma|^{1/2}} \left(1 + \frac{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+d}{2}},$$

  where $\boldsymbol{\mu} \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d\times d}$ is a positive definite matrix, and ν is the degrees of freedom. Notation: $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$.

- $t_d(\nu, \boldsymbol{\mu}, \Sigma)$ has heavier marginal and joint tails than $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$.

- BMW–Siemens data; simulations from fitted $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ and $t_d(3, \boldsymbol{\mu}, \Sigma)$:



6.2.2 Normal mean-variance mixtures

- Radial symmetry implies that all one-dimensional margins of normal variance mixtures are symmetric.
- Often visible in data: joint losses have heavier tails than joint gains.

Idea: Introduce asymmetry by mixing normal distributions with different means and variances.

$\boldsymbol{X}$ has a (multivariate) *normal mean-variance mixture distribution* if

$$\boldsymbol{X} \overset{\mathrm{d}}{=} \boldsymbol{m}(W) + \sqrt{W} A \boldsymbol{Z}, \tag{16}$$

where

- $\boldsymbol{Z} \sim \mathrm{N}_k(\boldsymbol{0}, I_k)$;
- $W \geq 0$ is a scalar random variable which is independent of $\boldsymbol{Z}$;
- $A \in \mathbb{R}^{d \times k}$ is a matrix of constants;
- $\boldsymbol{m} : [0, \infty) \to \mathbb{R}^d$ is a measurable function.

- Normal mean-variance mixtures add skewness: Let $\Sigma = AA'$ and observe that $\boldsymbol{X} \,|\, W = w \sim \mathrm{N}_d(\boldsymbol{m}(w), w\Sigma)$. In general, they are no longer elliptical (see later).

Example 6.9

- Suppose we have $\boldsymbol{m}(W) = \boldsymbol{\mu} + W\boldsymbol{\gamma}$. Since

$$\mathbb{E}(\boldsymbol{X} \,|\, W) = \boldsymbol{\mu} + W\boldsymbol{\gamma},$$
$$\mathrm{cov}(\boldsymbol{X} \,|\, W) = W\Sigma$$

we have

$$\begin{aligned}
\mathbb{E}\boldsymbol{X} &= \mathbb{E}(\mathbb{E}(\boldsymbol{X} \,|\, W)) = \boldsymbol{\mu} + \mathbb{E}(W)\boldsymbol{\gamma} \quad \text{if } \mathbb{E}W < \infty, \\
\mathrm{cov}(\boldsymbol{X}) &= \mathbb{E}(\mathrm{cov}(\boldsymbol{X} \,|\, W)) + \mathrm{cov}(\mathbb{E}(\boldsymbol{X} \,|\, W)) \\
&= \mathbb{E}(W)\Sigma + \mathrm{var}(W)\boldsymbol{\gamma}\boldsymbol{\gamma}' \quad \text{if } \mathbb{E}(W^2) < \infty.
\end{aligned}$$

- If W has a GIG distribution, then $\boldsymbol{X}$ follows a *generalised hyperbolic distribution*. $\boldsymbol{\gamma} = \mathbf{0}$ leads to (elliptical) normal variance mixtures; see MFE (2015, Sections 6.2.3) for details.

6.3 Spherical and elliptical distributions

Empirical examples (see MFE (2015, Sections 6.2.4)) show that

1) $M_d(\boldsymbol{\mu}, \Sigma, \hat{F}_W)$ (e.g. multivariate t) provide superior models to $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ for daily/weekly stock-return data;
2) the more general skewed normal mean-variance mixture distributions offer only a modest improvement.

We study elliptical distributions, a generalization of $M_d(\boldsymbol{\mu}, \Sigma, \hat{F}_W)$.

6.3.1 Spherical distributions

Definition 6.10 (Spherical distribution)
A random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ has a *spherical distribution* if for every orthogonal $U \in \mathbb{R}^{d \times d}$ (i.e. $U \in \mathbb{R}^{d \times d}$ with $UU' = U'U = I_d$)

$$\boldsymbol{Y} \stackrel{\mathrm{d}}{=} U\boldsymbol{Y} \quad \text{(distributionally invariant under rotations and reflections)}$$

Theorem 6.11 (Characterization of spherical distributions)
Let $\|\boldsymbol{t}\| = (t_1^2 + \cdots + t_d^2)^{1/2}$, $\boldsymbol{t} \in \mathbb{R}^d$. The following are equivalent:

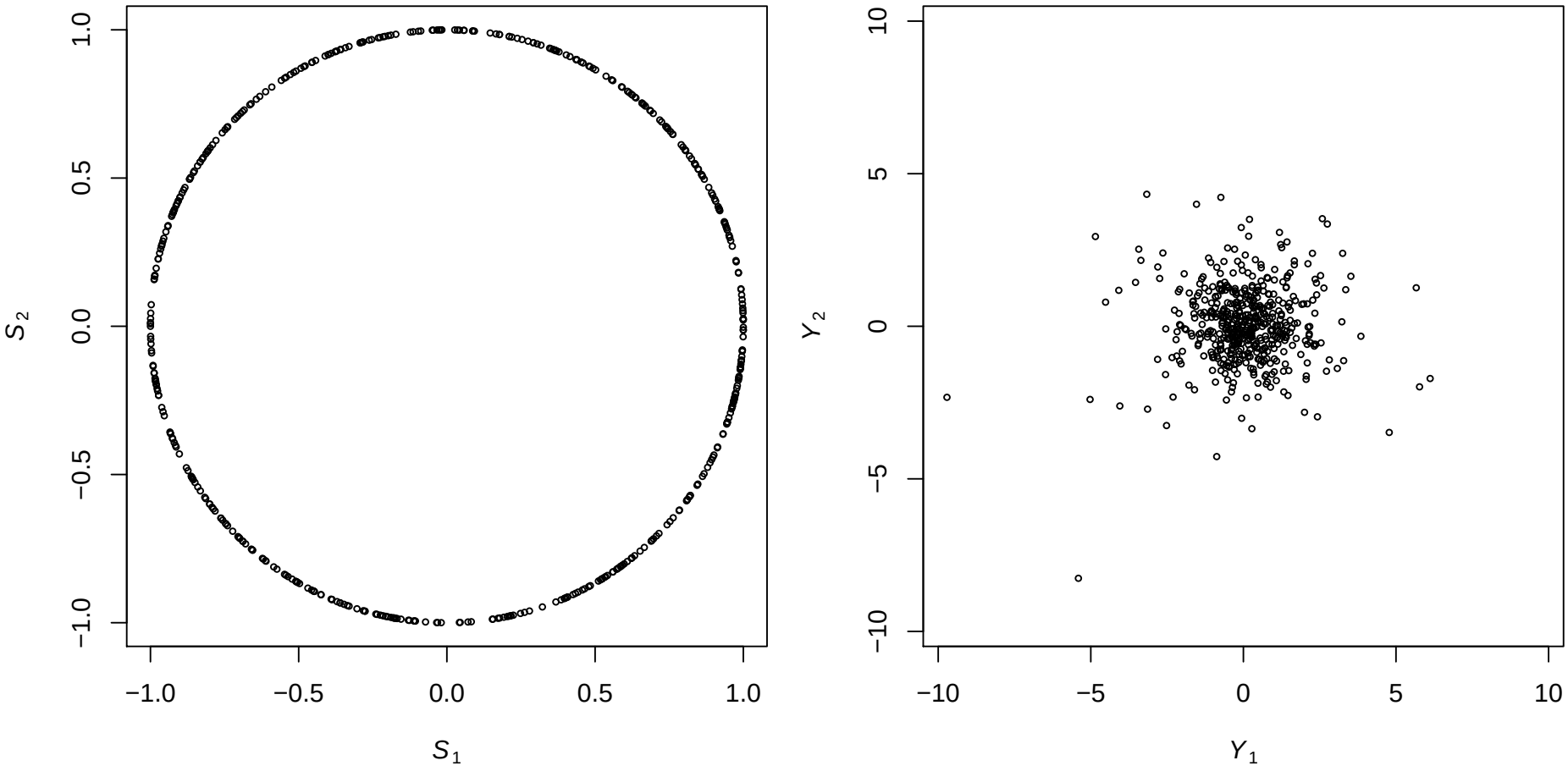
1) $\boldsymbol{Y}$ is spherical (notation: $\boldsymbol{Y} \sim S_d(\psi)$ for ψ as below).
2) $\exists$ a *characteristic generator* $\psi : [0, \infty) \to \mathbb{R}$, such that $\phi_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathbb{E}(e^{i\boldsymbol{t}'\boldsymbol{Y}}) = \psi(\|\boldsymbol{t}\|^2)$, $\forall\, \boldsymbol{t} \in \mathbb{R}^d$.
3) For every $\boldsymbol{a} \in \mathbb{R}^d$, $\boldsymbol{a}'\boldsymbol{Y} \overset{\mathrm{d}}{=} \|\boldsymbol{a}\| Y_1$ (lin. comb. are of the same type).
   $\Rightarrow$ Subadditivity of VaR_α for jointly elliptical losses

Theorem 6.12 (Stochastic representation)
$\boldsymbol{Y} \sim S_d(\psi)$ if and only if $\boldsymbol{Y} \overset{\mathrm{d}}{=} R\boldsymbol{S}$ for an independent *radial part* $R \geq 0$ and $\boldsymbol{S} \sim \mathrm{U}(\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| = 1\})$.

This is the key to understanding the structure of spherical distributions.

Example 6.13 (Understanding spherical distributions)

$n = 500$ realizations of $\boldsymbol{S}$ (left) and $\boldsymbol{Y} = R\boldsymbol{S}$ (right) for $R \sim \sqrt{dF(d, \nu)}$, $d = 2$, $\nu = 4$ (as for the multivariate t distribution with $\nu = 4$).



6.3.2 Elliptical distributions

Definition 6.14 (Elliptical distribution)
A random vector $\boldsymbol{X} = (X_1, \dots, X_d)$ has an *elliptical distribution* if

$$\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + A\boldsymbol{Y}, \quad \text{(multivariate affine transformation)}$$

where $\boldsymbol{Y} \sim S_k(\psi)$, $A \in \mathbb{R}^{d \times k}$ (*scale matrix* $\Sigma = AA'$), and (*location vector*) $\boldsymbol{\mu} \in \mathbb{R}^d$.

- By Theorem 6.12, an elliptical random vector admits the stochastic representation $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + RA\boldsymbol{S}$, with R and $\boldsymbol{S}$ as before.
- Notation: $\boldsymbol{X} \sim \mathrm{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$
- If $\boldsymbol{X} \sim \mathrm{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$ with $\mathbb{P}(\boldsymbol{X} = \boldsymbol{\mu}) = 0$, then $\boldsymbol{Y} = A^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim S_d(\psi)$. One can then show that

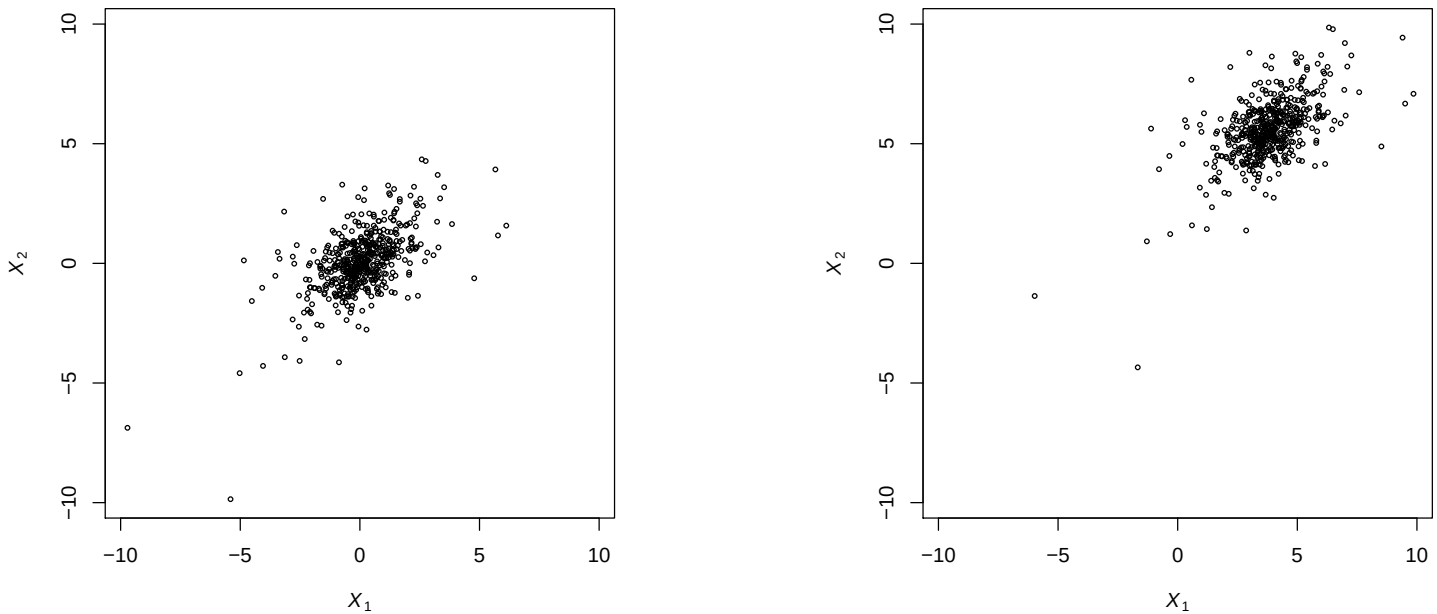
$$\left(\sqrt{(\boldsymbol{X} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{X} - \boldsymbol{\mu})}, \ \frac{A^{-1}(\boldsymbol{X} - \boldsymbol{\mu})}{\sqrt{(\boldsymbol{X} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{X} - \boldsymbol{\mu})}} \right) \stackrel{\mathrm{d}}{=} (R, \boldsymbol{S}), \qquad (17)$$

which can be used for testing elliptical symmetry.

- Normal variance mixture distributions are elliptical (most useful examples) since $\boldsymbol{X} \overset{\text{d}}{=} \boldsymbol{\mu} + \sqrt{W} A \boldsymbol{Z} = \boldsymbol{\mu} + \sqrt{W} \|\boldsymbol{Z}\| A \boldsymbol{Z}/\|\boldsymbol{Z}\| = \boldsymbol{\mu} + RA\boldsymbol{S}$ with $R = \sqrt{W}\|\boldsymbol{Z}\|$ and $\boldsymbol{S} = \boldsymbol{Z}/\|\boldsymbol{Z}\|$ (independent).

Example 6.15 (Understanding elliptical distributions)

$n = 500$ realizations of $\boldsymbol{X} = RA\boldsymbol{S}$ (left) and $\boldsymbol{X} = \boldsymbol{\mu} + RA\boldsymbol{S}$ (right) for $R \sim \sqrt{dF(d,\nu)}$, $d = 2$, $\nu = 4$; recycling of samples from Example 6.13.



6.3.3 Properties of elliptical distributions

- **Density:** Let Σ be positive definite and $\boldsymbol{Y} \sim S_d(\psi)$ have density generator g. The density transformation theorem implies that $\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Y}$ has density

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{\det \Sigma}} g((\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})),$$

  which depends on $\boldsymbol{x}$ only through $(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$, i.e. is constant on ellipsoids (hence the name "elliptical").

- **Linear combinations:** For $\boldsymbol{X} \sim \mathrm{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$, $B \in \mathbb{R}^{k\times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$,

$$B\boldsymbol{X} + \boldsymbol{b} \sim \mathrm{E}_k(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B', \psi) \quad \text{(via cfs)}.$$

  If $\boldsymbol{a} \in \mathbb{R}^d$ (take $\boldsymbol{b} = \boldsymbol{0}$ and $B = \boldsymbol{a}' \in \mathbb{R}^{1\times d}$),

$$\boldsymbol{a}'\boldsymbol{X} \sim \mathrm{E}_1(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\Sigma\boldsymbol{a}, \psi) \quad \text{(as for } \mathrm{N}(\boldsymbol{\mu}, \Sigma)\text{)}. \tag{18}$$

  From $\boldsymbol{a} = \boldsymbol{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ we see that all marginal distributions are of the same type.

- **Marginal dfs:** As for $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$, it immediately follows that $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')' \sim \mathrm{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$ satisfies $\boldsymbol{X}_1 \sim \mathrm{E}_k(\boldsymbol{\mu}_1, \Sigma_{11}, \psi)$ and that $\boldsymbol{X}_2 \sim \mathrm{E}_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22}, \psi)$; i.e. margins of elliptical distributions are elliptical.
- **Quadratic forms:** $(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \overset{\mathrm{d}}{=} R^2$. If $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$, $R^2 \sim \chi^2_d$; and if $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$, $R^2/d \sim F(d, \nu)$.

Many (but not all) nice properties of $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ are preserved. The following result shows why elliptical distributions are the "Garden of Eden" of QRM.

Proposition 6.16 (Subadditivity of VaR in elliptical models) Let $L_i = \boldsymbol{\lambda}_i' \boldsymbol{X}$, $\boldsymbol{\lambda}_i \in \mathbb{R}^d$, $i \in \{1, \dots, n\}$, with $\boldsymbol{X} \sim \mathrm{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$. Then $\mathrm{VaR}_\alpha(\sum_{i=1}^n L_i) \leq \sum_{i=1}^n \mathrm{VaR}_\alpha(L_i)$ for all $\alpha \in [1/2, 1]$.

Proof. Consider a generic $L = \boldsymbol{\lambda}' \boldsymbol{X} \overset{\mathrm{d}}{=} \boldsymbol{\lambda}' \boldsymbol{\mu} + \boldsymbol{\lambda}' A \boldsymbol{Y}$ for $\boldsymbol{Y} \sim S_k(\psi)$. By Theorem 6.11 Part 3), $\boldsymbol{\lambda}' A \boldsymbol{Y} \overset{\mathrm{d}}{=} \|\boldsymbol{\lambda}' A\| Y_1$, so $L \overset{\mathrm{d}}{=} \boldsymbol{\lambda}' \boldsymbol{\mu} + \|\boldsymbol{\lambda}' A\| Y_1$ (all L_i's are of the same type). By translation invariance and positive homogeneity,

$$\mathrm{VaR}_\alpha(L) = \boldsymbol{\lambda}' \boldsymbol{\mu} + \|\boldsymbol{\lambda}' A\| \, \mathrm{VaR}_\alpha(Y_1). \tag{19}$$

Applying (19) once to $L = \sum_{i=1}^n L_i = (\sum_{i=1}^n \boldsymbol{\lambda}_i)' \boldsymbol{X}$ and to each $L = L_i = \boldsymbol{\lambda}_i' \boldsymbol{X}$, $i \in \{1, \dots, n\}$, and using that $\mathrm{VaR}_\alpha(Y_1) \geq 0$ for $\alpha \in [1/2, 1]$, we obtain $\mathrm{VaR}_\alpha(\sum_{i=1}^n L_i) \underset{(19)}{=} \sum_{i=1}^n \boldsymbol{\lambda}_i' \boldsymbol{\mu} + \|\sum_{i=1}^n \boldsymbol{\lambda}_i' A\| \, \mathrm{VaR}_\alpha(Y_1)$
$\leq \sum_{i=1}^n \boldsymbol{\lambda}_i' \boldsymbol{\mu} + (\sum_{i=1}^n \|\boldsymbol{\lambda}_i' A\|) \, \mathrm{VaR}_\alpha(Y_1) = \sum_{i=1}^n (\boldsymbol{\lambda}_i' \boldsymbol{\mu} + \|\boldsymbol{\lambda}_i' A\| \, \mathrm{VaR}_\alpha(Y_1))$
$\underset{(19)}{=} \sum_{i=1}^n \mathrm{VaR}_\alpha(L_i)$. For $\boldsymbol{\lambda}_i = \boldsymbol{e}_i$, $\mathrm{VaR}_\alpha(\sum_{i=1}^n X_i) \leq \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)$. □

7 Copulas and dependence

7.1 Copulas

7.2 Dependence concepts and measures

7.3 A proof for subadditivity of ES

7.4 Fitting copulas to data

7.1 Copulas

- We now look more closely at modelling the dependence among the components of a random vector $\boldsymbol{X} \sim F$ (risk-factor changes).
- **In short:** $F =$ dependence structure $C \circ$ marginal dfs $F_1, \ldots, F_d$
- **Advantages:**
  - Most natural in a static distributional context (no time dependence; apply, for example, to residuals of an ARMA-GARCH model)
  - Copulas allow us to understand and study dependence independently of the margins (first part of Sklar's Theorem; see later)
  - Copulas allow for a bottom-up approach to multivariate model building (second part of Sklar's Theorem; see later). This is often useful for constructing tailored F, for example, when we have more information about the margins than C or for *stress testing* purposes (to challenge the existing model and see how it performs).

7.1.1 Basic properties

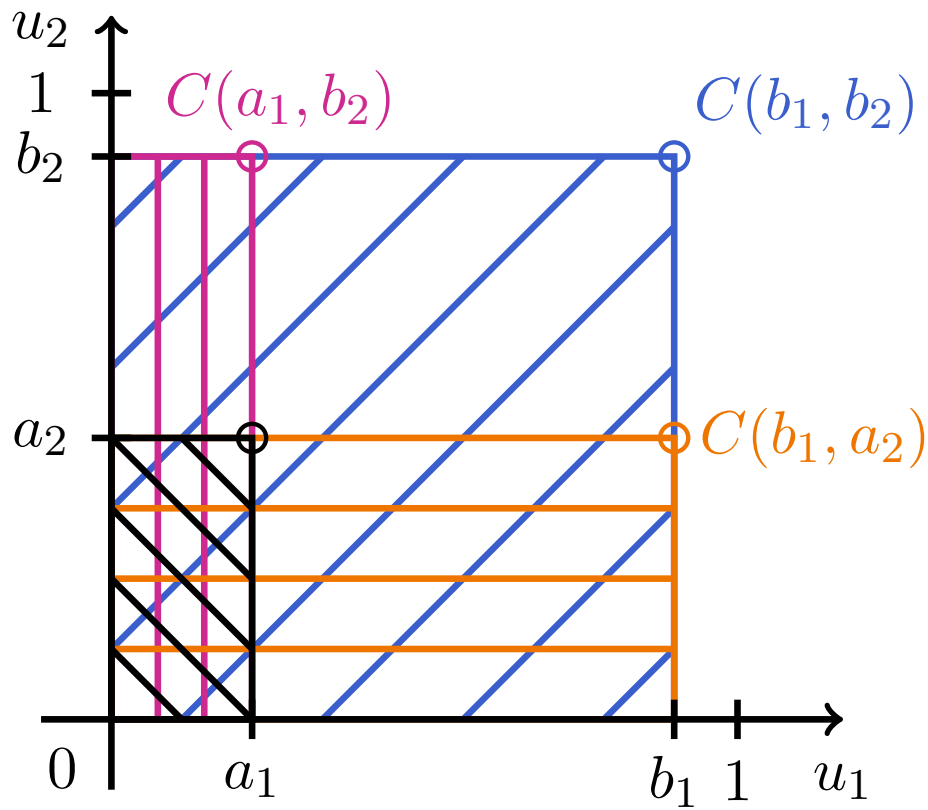
Definition 7.1 (Copula)
A *copula* C is a df with $\mathrm{U}(0,1)$ margins.

Characterization
$C : [0,1]^d \to [0,1]$ is a copula if and only if

1) C is *grounded*, that is,
   $C(u_1, \dots, u_d) = 0$ if $u_j = 0$ for at least one $j \in \{1, \dots, d\}$.
2) C has standard *uniform* univariate *margins*, that is,
   $C(1, \dots, 1, u_j, 1, \dots, 1) = u_j$ for all $u_j \in [0,1]$ and $j \in \{1, \dots, d\}$.
3) C is *d-increasing*, that is, for all $\boldsymbol{a}, \boldsymbol{b} \in [0,1]^d$, $\boldsymbol{a} \leq \boldsymbol{b}$, $\Delta_{(\boldsymbol{a},\boldsymbol{b}]} C = \sum_{\boldsymbol{i} \in \{0,1\}^d} (-1)^{\sum_{j=1}^d i_j} C(a_1^{i_1} b_1^{1-i_1}, \dots, a_d^{i_d} b_d^{1-i_d}) \geq 0$. Equivalently, if it exists, the density c of C satisfies $c(\boldsymbol{u}) \geq 0$ for all $\boldsymbol{u} \in (0,1)^d$.

2-increasingness explained in a picture:

$$\begin{aligned}\Delta_{(\boldsymbol{a},\boldsymbol{b}]}C &= C(b_1,b_2) - C(b_1,a_2) - C(a_1,b_2) + C(a_1,a_2)\\ &= \mathbb{P}(\boldsymbol{U} \in (\boldsymbol{a},\boldsymbol{b}]) \overset{!}{\geq} 0\end{aligned}$$



$\Rightarrow \Delta_{(\boldsymbol{a},\boldsymbol{b}]}C$ is the probability of a random vector $\boldsymbol{U} \sim C$ to be in $(\boldsymbol{a},\boldsymbol{b}]$.

Preliminaries

Lemma 7.2 (Probability transformation)
Let $X \sim F$, F continuous. Then $F(X) \sim \mathrm{U}(0,1)$.

Proof. $\mathbb{P}(F(X) \leq u) = \mathbb{P}(F^{\leftarrow}(F(X)) \leq F^{\leftarrow}(u)) = \mathbb{P}(X \leq F^{\leftarrow}(u)) = F(F^{\leftarrow}(u)) = u$, $u \in [0,1]$. □

Note that F needs to be continuous (otherwise $F(X)$ would not reach all intervals $\subseteq [0,1]$).

Lemma 7.3 (Quantile transformation)
Let $U \sim \mathrm{U}(0,1)$ and F be any df. Then $X = F^{\leftarrow}(U) \sim F$.

Proof. $\mathbb{P}(F^{\leftarrow}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$, $x \in \mathbb{R}$. □

Probability and quantile transformations are the key to all applications involving copulas. They allow us to go from $\mathbb{R}^d$ to $[0,1]^d$ and back.

Sklar's Theorem

Theorem 7.4 (Sklar's Theorem)

1) For any df F with margins $F_1, \dots, F_d$, there exists a copula C such that

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)), \quad \boldsymbol{x} \in \mathbb{R}^d. \tag{20}$$

C is uniquely defined on $\prod_{j=1}^{d} \operatorname{ran} F_j$ and given by

$$C(u_1, \dots, u_d) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)), \quad \boldsymbol{u} \in \prod_{j=1}^{d} \operatorname{ran} F_j,$$

where $\operatorname{ran} F_j = \{F_j(x) : x \in \mathbb{R}\}$ denotes the *range* of F_j.

2) Conversely, given any copula C and univariate dfs $F_1, \dots, F_d$, F defined by (20) is a df with margins $F_1, \dots, F_d$.

Proof.

1) Proof for continuous $F_1, \dots, F_d$ only. Let $\boldsymbol{X} \sim F$ and define $U_j = F_j(X_j)$, $j \in \{1, \dots, d\}$. By the probability transformation, $U_j \sim \mathrm{U}(0,1)$ (continuity!), $j \in \{1, \dots, d\}$, so the df C of $\boldsymbol{U}$ is a copula. Since $F_j \uparrow$ on $\operatorname{ran} X_j$, $X_j = F_j^{\leftarrow}(F_j(X_j)) = F_j^{\leftarrow}(U_j)$, $j \in \{1, \dots, d\}$. Therefore,

$$F(\boldsymbol{x}) = \mathbb{P}(X_j \le x_j \ \forall j) = \mathbb{P}(F_j^{\leftarrow}(U_j) \le x_j \ \forall j) = \mathbb{P}(U_j \le F_j(x_j) \ \forall j)$$
$$= C(F_1(x_1), \dots, F_d(x_d)), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

Hence C is a copula and satisfies (20).

Since $F_j(F_j^{\leftarrow}(u_j)) = u_j$ for all $u_j \in \operatorname{ran} F_j$, so

$$C(u_1, \dots, u_d) = C(F_1(F_1^{\leftarrow}(u_1)), \dots, F_d(F_d^{\leftarrow}(u_d)))$$
$$\underset{(20)}{=} F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)), \quad \boldsymbol{u} \in \prod_{j=1}^{d} \operatorname{ran} F_j.$$

2) For $\boldsymbol{U} \sim C$, define $\boldsymbol{X} = (F_1^{\leftarrow}(U_1), \dots, F_d^{\leftarrow}(U_d))$. Then

$$\mathbb{P}(\boldsymbol{X} \le \boldsymbol{x}) = \mathbb{P}(F_j^{\leftarrow}(U_j) \le x_j \ \forall\, j) = \mathbb{P}(U_j \le F_j(x_j) \ \forall\, j)$$
$$= C(F_1(x_1), \dots, F_d(x_d)), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

Therefore, F defined by (20) is a df (that of $\boldsymbol{X}$), with margins $F_1, \dots, F_d$ (obtained by the quantile transformation). □

- We say that $\boldsymbol{X}$ (or F) *has copula* C if (20) holds.
- A *copula model* for $\boldsymbol{X}$ means $F(\boldsymbol{x}) = C(F_1(x_1), \dots, F_d(x_d))$ for some (parametric) copula C and (parametric) marginals $F_1, \dots, F_d$.

Invariance principle

Theorem 7.5 (Invariance principle)
Let $\boldsymbol{X} \sim F$ with continuous margins $F_1, \dots, F_d$ and copula C. If $T_j \uparrow$ on $\operatorname{ran} X_j$ for all j, then $(T_1(X_1), \dots, T_d(X_d))$ (also) has copula C.

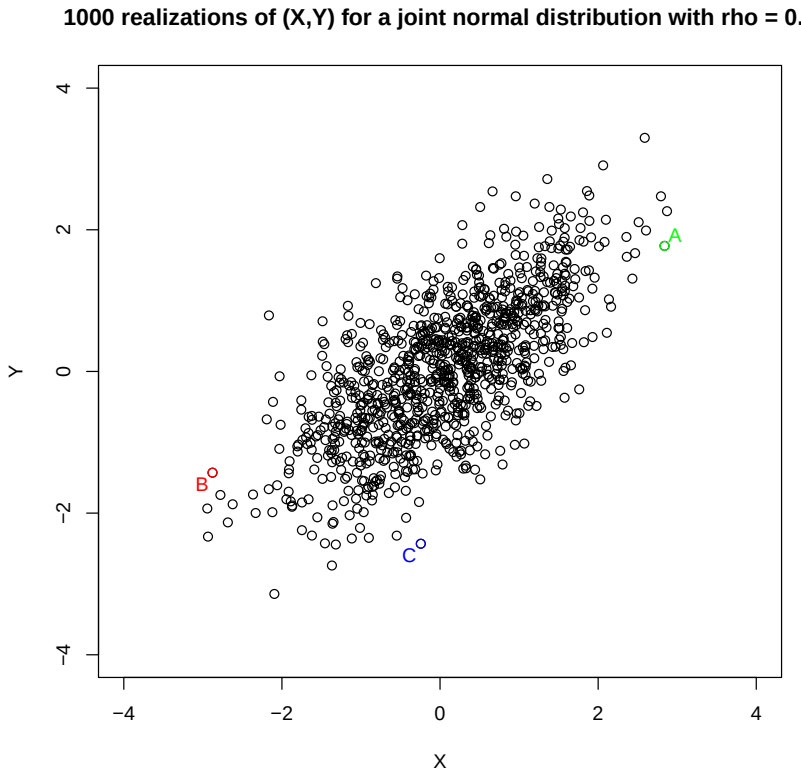
Interpretation of Sklar's Theorem (and the invariance principle)

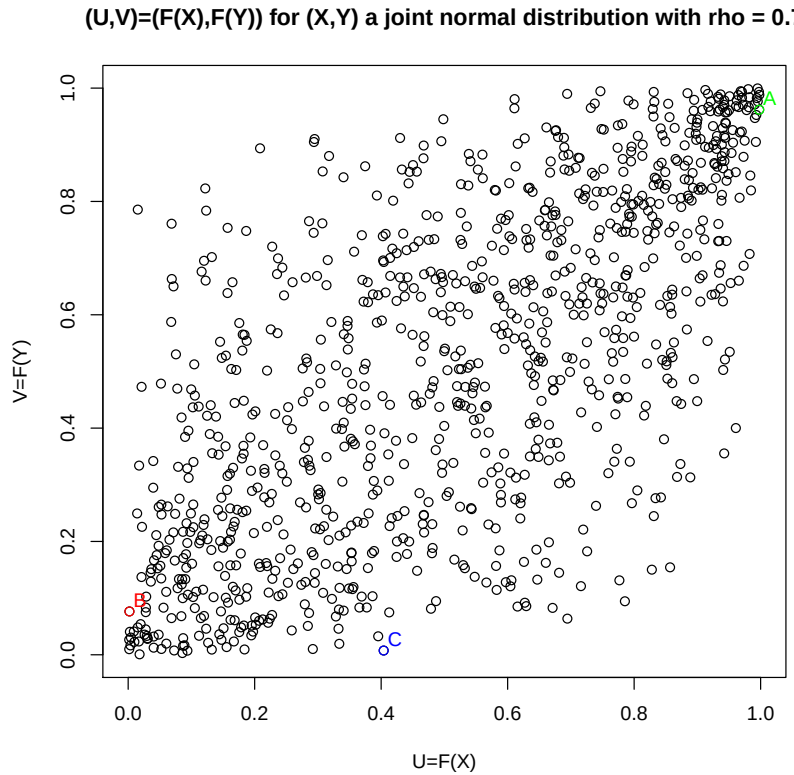
1) Part 1) of Sklar's Theorem allows one to decompose any df F into its margins and a copula. This, together with the invariance principle, allows one to study dependence independently of the margins via the margin-free $\boldsymbol{U} = (F_1(X_1), \dots, F_d(X_d))$ instead of $\boldsymbol{X} = (X_1, \dots, X_d)$ (they both have the same copula!). This is interesting for statistical applications, e.g. parameter estimation or goodness-of-fit.
2) Part 2) allows one to construct flexible multivariate distributions for particular applications (credit risk, stress testing, etc.).

Visualizing Part 1) of Sklar's Theorem

Left: Scatter plot of $n = 1000$ samples from $(X_1, X_2) \sim \mathrm{N}_2(\mathbf{0}, P)$, where $P = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$. We mark three points A, B, C.

Right: Scatter plot of the corresponding Gauss copula (after applying the df Φ of $\mathrm{N}(0,1)$). Note how A, B, C change.

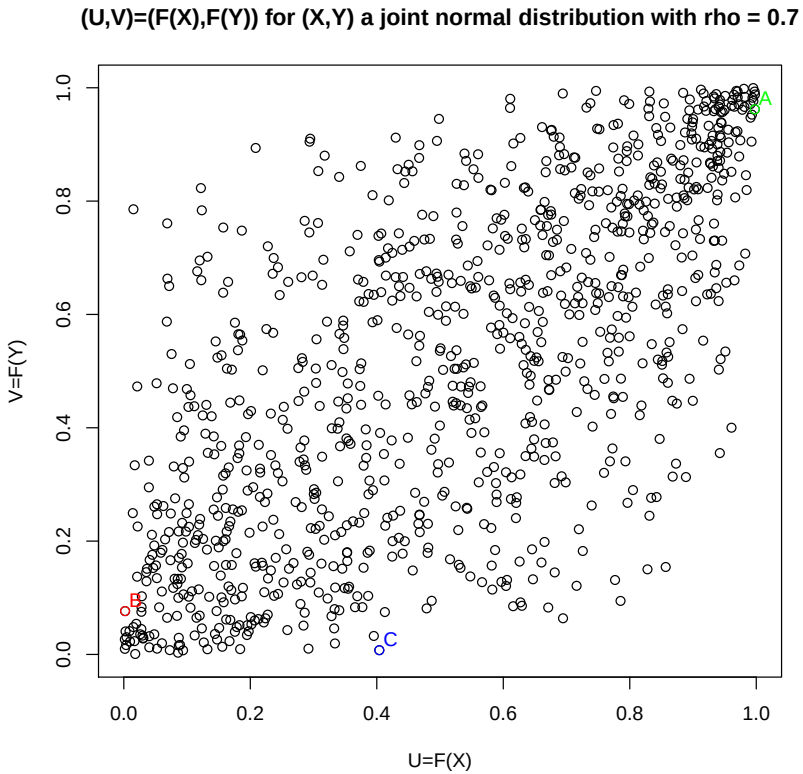


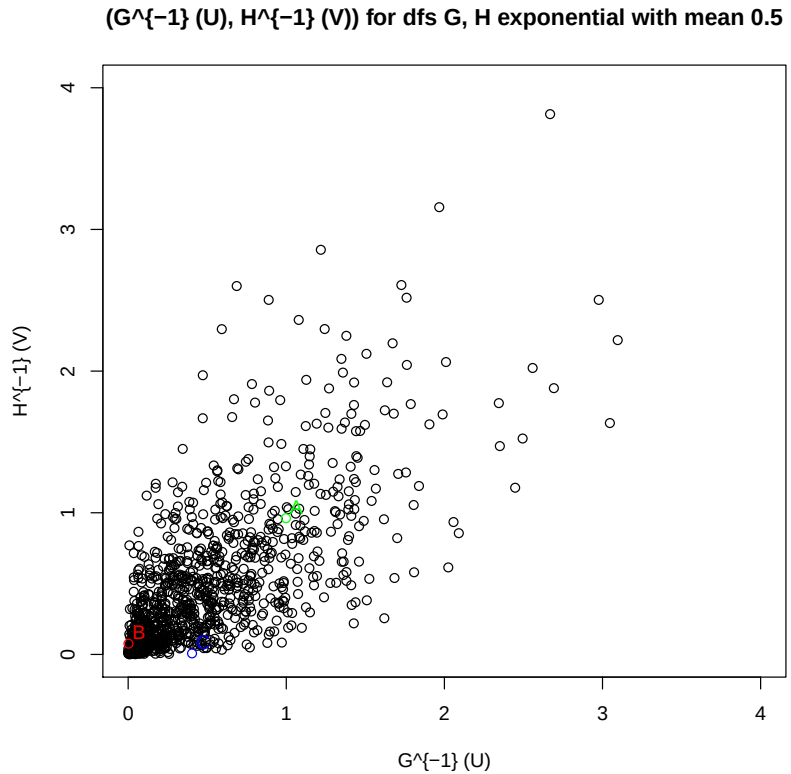


Visualizing Part 2) of Sklar's Theorem

Left: Same Gauss copula scatter plot as before. Apply marginal $\mathrm{Exp}(2)$-quantile functions $(F_j^{-1}(u) = -\log(1-u)/2,\ j \in \{1,2\})$.

Right: The corresponding transformed random variates. Again, note the three points A, B, C.

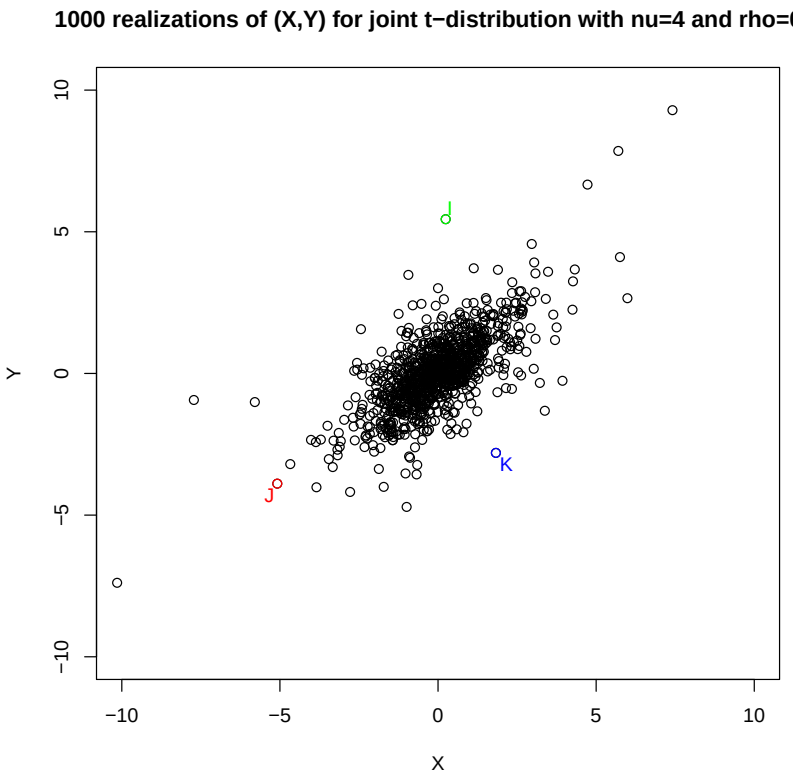


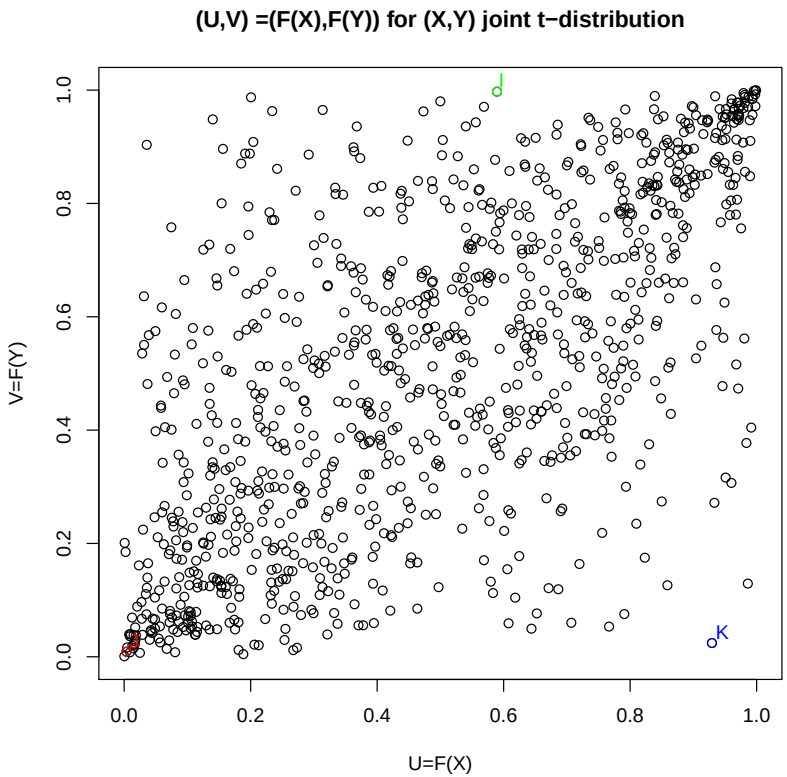


Visualizing Part 1) of Sklar's Theorem

Left: Scatter plot of $n = 1000$ samples from $(X_1, X_2) \sim t_2(4, \mathbf{0}, P)$, where $P = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$. We mark three points I, J, K.

Right: Scatter plot of the corresponding t_4 copula (after applying the df t_4). Note how I, J, K change.

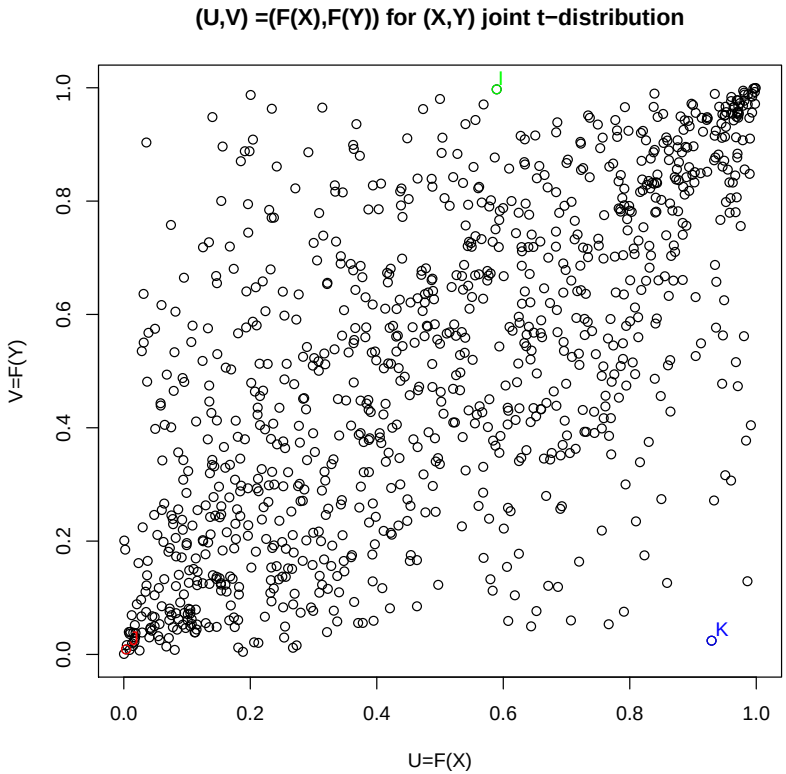


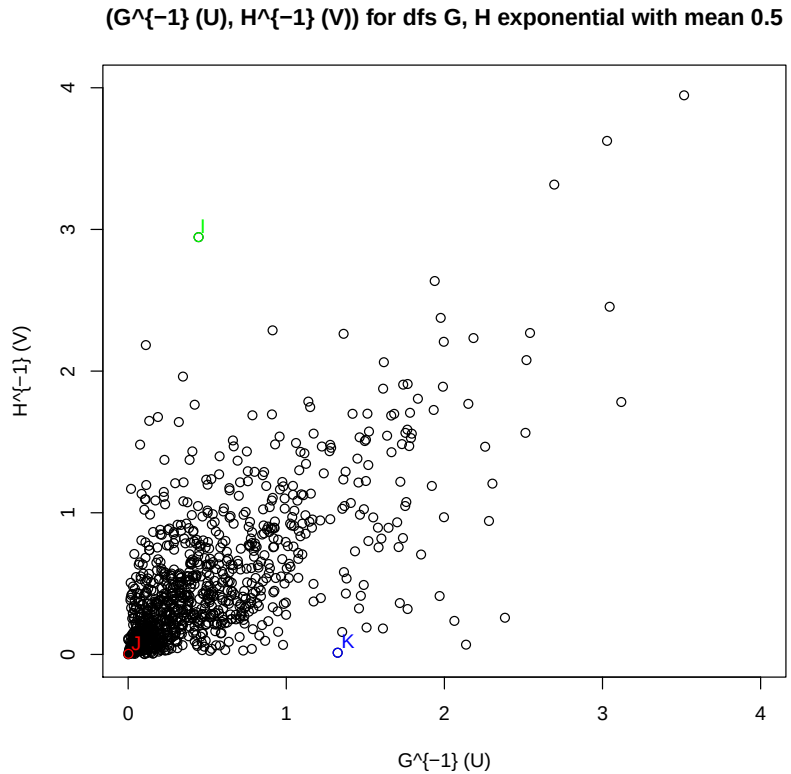


Visualizing Part 2) of Sklar's Theorem

Left: Same t_4 copula scatter plot as before. Apply marginal $\mathrm{Exp}(2)$-quantile functions $(F_j^{-1}(u) = -\log(1-u)/2,\ j \in \{1,2\})$.

Right: The corresponding transformed random variates. Again, note the three points I, J, K.





Fréchet–Hoeffding bounds

Theorem 7.6 (Fréchet–Hoeffding bounds)
Let $W(\boldsymbol{u}) = \max\{\sum_{j=1}^{d} u_j - d + 1, 0\}$ and $M(\boldsymbol{u}) = \min_{1 \le j \le d}\{u_j\}$.

1) For any d-dimensional copula C,

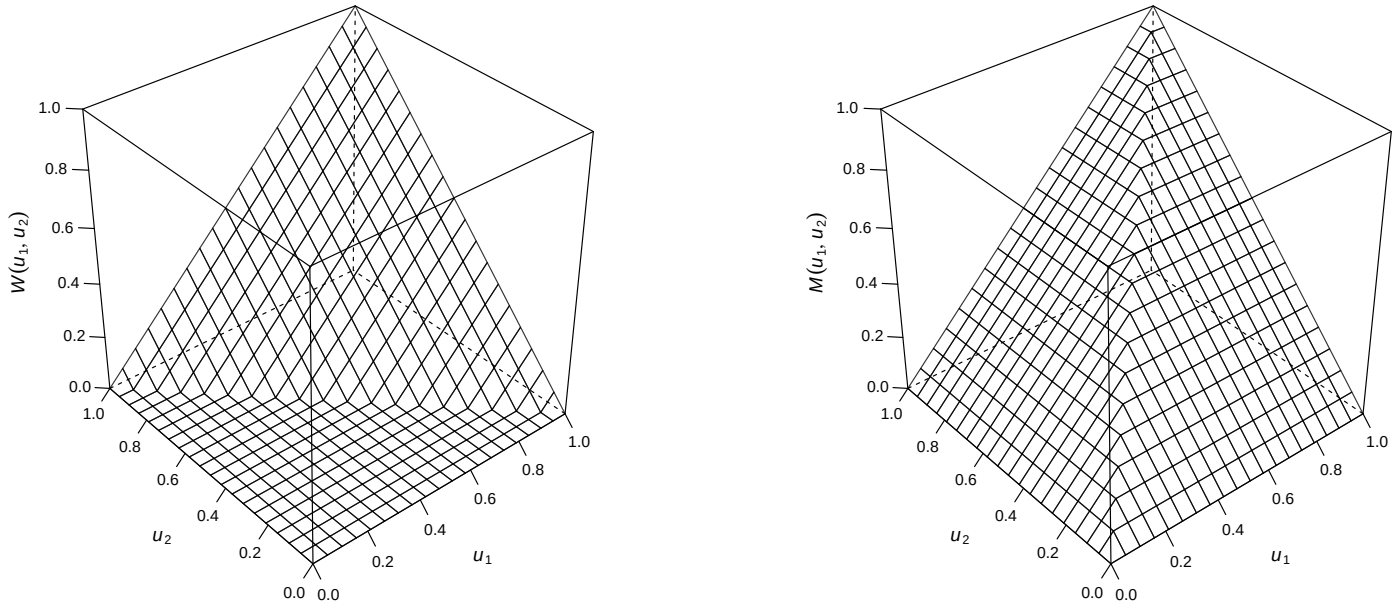
$$W(\boldsymbol{u}) \le C(\boldsymbol{u}) \le M(\boldsymbol{u}), \quad \boldsymbol{u} \in [0,1]^d.$$

2) W is a copula if and only if $d = 2$.

3) M is a copula for all $d \ge 2$.

- It is easy to verify that, for $U \sim \mathrm{U}(0,1)$,
  - $(U, \dots, U) \sim M$;
  - $(U, 1-U) \sim W$.

- Plot of W, M for $d = 2$ (compare with $(U, 1-U) \sim W$, $(U, U) \sim M$)



- The Fréchet–Hoeffding bounds correspond to perfect dependence (negative for W; positive for M); see Proposition 7.10 later.
- The Fréchet–Hoeffding bounds lead to bounds for any df F, via

$$\max\Big\{\sum_{j=1}^{d} F_j(x_j) - d + 1, 0\Big\} \le F(\boldsymbol{x}) \le \min_{1\le j\le d}\{F_j(x_j)\}.$$

We will use them later to derive bounds for the correlation coefficient.

7.1.2 Examples of copulas

- *Fundamental copulas*: important special copulas;
- *Implicit copulas*: extracted from known F via Sklar's Theorem;
- *Explicit copulas*: closed form, arising from construction principles.

Fundamental copulas

As usual, we assume the appearing margins $F_1, \dots, F_d$ to be continuous.

- $\Pi(\boldsymbol{u}) = \prod_{j=1}^d u_j$ is the *independence copula* since $C(F_1(x_1), \dots, F_d(x_d)) \underset{\text{Sklar}}{=} F(\boldsymbol{x}) \underset{\text{ind.}}{=} \prod_{j=1}^d F_j(x_j)$ if and only if $C(\boldsymbol{u}) = \Pi(\boldsymbol{u})$ (replace x_j by $F_j^{\leftarrow}(u_j)$). Therefore, $X_1, \dots, X_d$ are independent if and only if their copula is Π; the density is thus $c(\boldsymbol{u}) = 1$, $\boldsymbol{u} \in [0,1]^d$.
- W is the *countermonotonicity copula*. It is the df of $(U, 1-U)$. It can be shown that if X_1, X_2 are perfectly negatively dependent (X_2 is a.s. a strictly decreasing function of X_1), their copula is W.

- M is the *comonotonicity copula*. It is the df of $(U, \dots, U)$. It can be shown that if $X_1, \dots, X_d$ are perfectly positively dependent ($X_2, \dots, X_d$ are a.s. strictly increasing functions of X_1), their copula is M.

Implicit copulas

Elliptical copulas are implicit copulas arising from elliptical distributions via Sklar's Theorem. The two most prominent parametric families are the Gauss copula and the t copula (stemming from normal variance mixtures).

Gauss copulas

- Consider (w.l.o.g.) $\boldsymbol{X} \sim \mathrm{N}_d(\mathbf{0}, P)$. The *Gauss copula* (family) is given by

$$\begin{aligned} C_P^{\mathsf{Ga}}(\boldsymbol{u}) &= \mathbb{P}(\Phi(X_1) \leq u_1, \dots, \Phi(X_d) \leq u_d) \\ &= \Phi_P(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)) \end{aligned}$$

  where Φ_P is the df of $\mathrm{N}_d(\mathbf{0}, P)$ and Φ the df of $\mathrm{N}(0,1)$.

- Special cases: If $P = I_d$ then $C = \Pi$, and if $P = J_d = \mathbf{1}\mathbf{1}'$ then $C = M$. If $d = 2$ and $\rho = P_{12} = -1$ then $C = W$.
- Sklar's Theorem $\Rightarrow$ The density of $C(\boldsymbol{u}) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d))$ is

$$c(\boldsymbol{u}) = \frac{f(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d))}{\prod_{j=1}^d f_j(F_j^{\leftarrow}(u_j))}, \quad \boldsymbol{u} \in (0,1)^d.$$

  In particular, the density of C_P^{Ga} is

$$c_P^{\mathsf{Ga}}(\boldsymbol{u}) = \frac{1}{\sqrt{\det P}} \exp\Big(-\frac{1}{2}\boldsymbol{x}'(P^{-1} - I_d)\boldsymbol{x}\Big), \tag{21}$$

  where $\boldsymbol{x} = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$.

t copulas

- Consider (w.l.o.g.) $\boldsymbol{X} \sim t_d(\nu, \mathbf{0}, P)$. The *$t$ copula* (family) is given by

$$\begin{aligned} C_{\nu,P}^t(\boldsymbol{u}) &= \mathbb{P}(t_\nu(X_1) \le u_1, \dots, t_\nu(X_d) \le u_d) \\ &= t_{\nu,P}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d)) \end{aligned}$$

where $t_{\nu,P}$ is the df of $t_d(\nu, \mathbf{0}, P)$ and t_ν the df of the univariate t distribution with ν degrees of freedom.

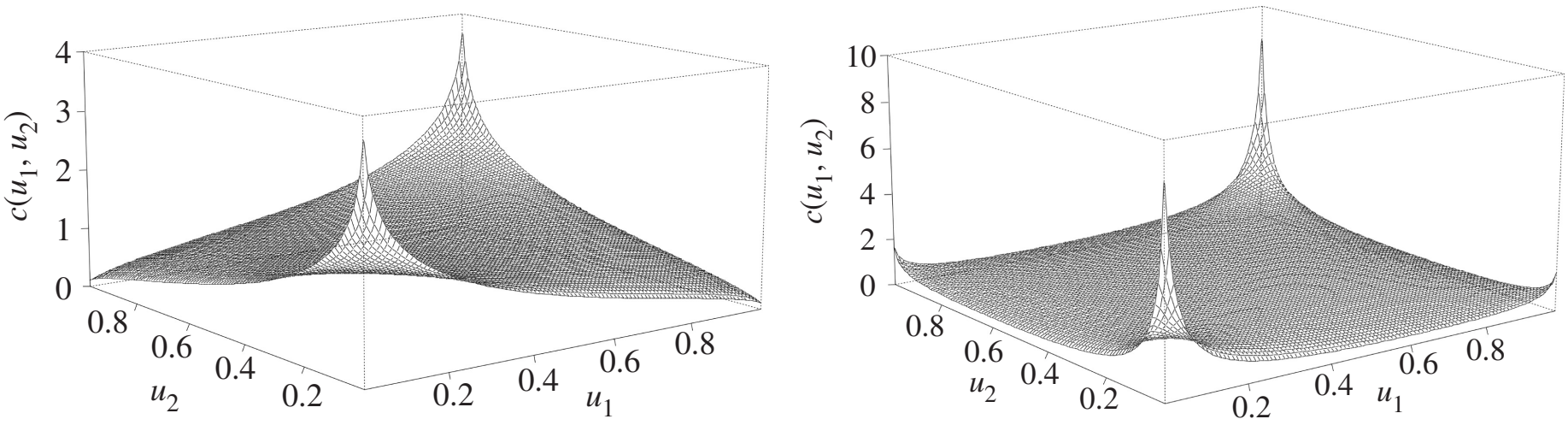
- Special cases: $P = J_d = \mathbf{1}\mathbf{1}'$ then $C = M$. However, if $P = I_d$ then $C \neq \Pi$ (unless $\nu = \infty$ in which case $C^t_{\nu,P} = C^{\text{Ga}}_P$). If $d = 2$ and $\rho = P_{12} = -1$ then $C = W$.
- Sklar's Theorem $\Rightarrow$ The density of $C^t_{\nu,P}$ is

$$c^t_{\nu,P}(\boldsymbol{u}) = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\sqrt{\det P}} \left(\frac{\Gamma(\nu/2)}{\Gamma((\nu+1)/2)} \right)^d \frac{(1 + \boldsymbol{x}' P^{-1} \boldsymbol{x}/\nu)^{-(\nu+d)/2}}{\prod_{j=1}^d (1 + x_j^2/\nu)^{-(\nu+1)/2}},$$

for $\boldsymbol{x} = (t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d))$.

- For more details, see Demarta and McNeil (2005).
- For scatter plots, see the visualization of Sklar's Theorem above. Note the difference in the tails: The smaller ν, the more mass is concentrated in the joint tails.

Perspective plots of the densities of $C^{\mathsf{Ga}}_{\rho=0.3}$ (left) and $C^{t}_{4,\,\rho=0.3}(\boldsymbol{u})$ (right).



Advantages and drawbacks of elliptical copulas:

Advantages:

- Modelling pairwise dependencies (comparably flexible)
- Density available
- Sampling simple (for Gauss, t)

Drawbacks:

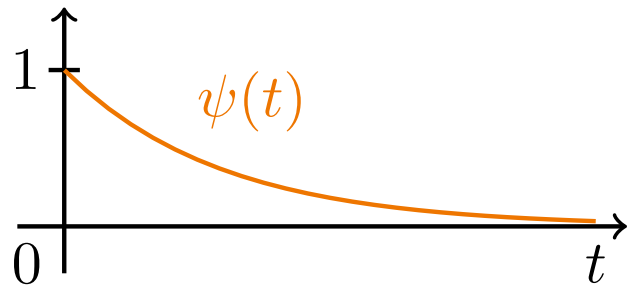
- Typically, C is not explicit
- Radially symmetric (so the same lower/upper tail behaviour)

Explicit copulas

Archimedean copulas are copulas of the form

$$C(\boldsymbol{u}) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$$

where ψ is the *(Archimedean) generator*.

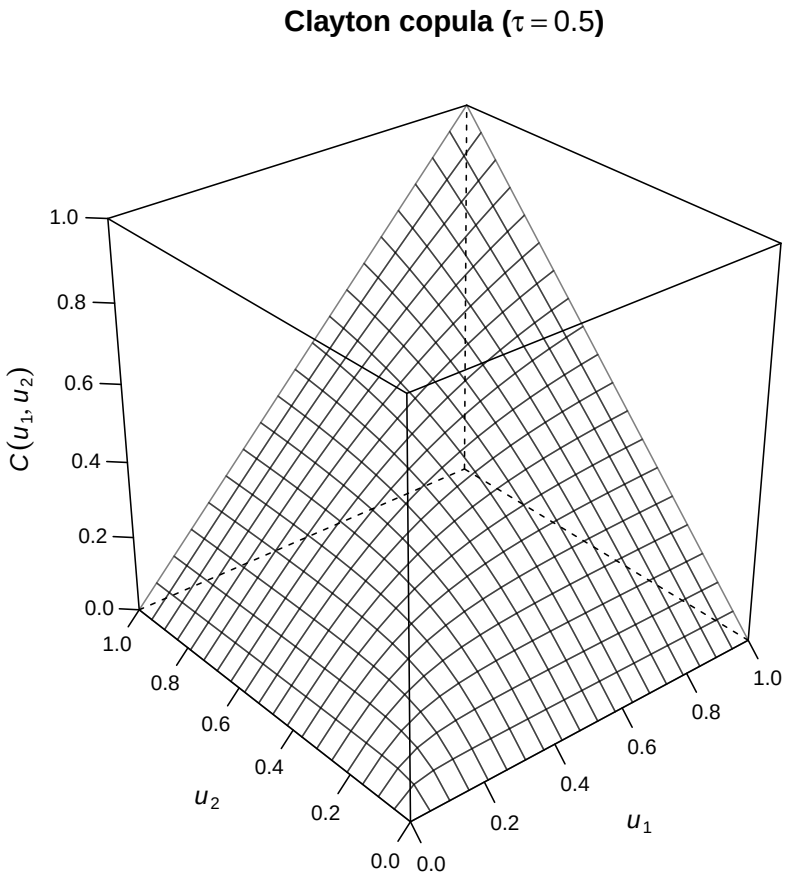


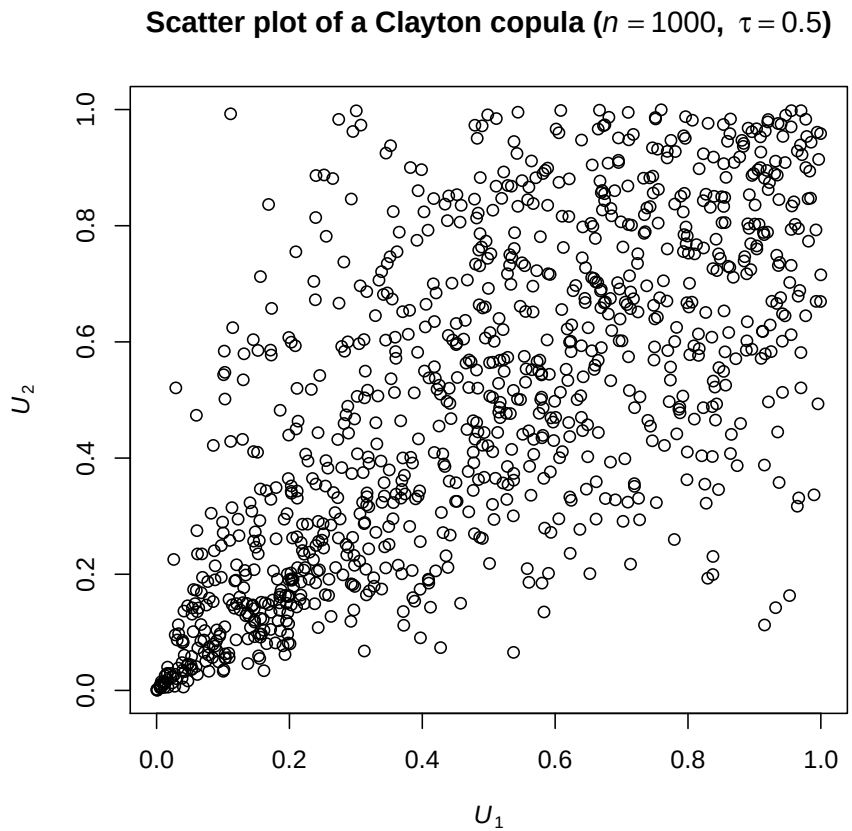
- $\psi : [0, \infty) \to [0, 1]$ is $\downarrow$ on $[0, \inf\{t : \psi(t) = 0\}]$ and satisfies $\psi(0) = 1$, $\psi(\infty) = \lim_{t\to\infty} \psi(t) = 0$.
- We set $\psi^{-1}(0) = \inf\{t : \psi(t) = 0\}$.
- The set of all generators is denoted by Ψ.
- Not every generator $\psi \in \Psi$ generates indeed a proper copula (there are conditions, e.g. complete monotonicity, i.e. derivatives alternating in sign).
- If $\psi(t) > 0$, $t \in [0, \infty)$, we call ψ *strict*.

Clayton copulas are obtained for $\psi(t) = (1+t)^{-1/\theta}$, $t \in [0, \infty)$, $\theta \in (0, \infty)$. For $\theta \downarrow 0$, $C \to \Pi$; and for $\theta \uparrow \infty$, $C \to M$.

Left: Plot of a bivariate Clayton copula (Kendall's tau 0.5; see later).

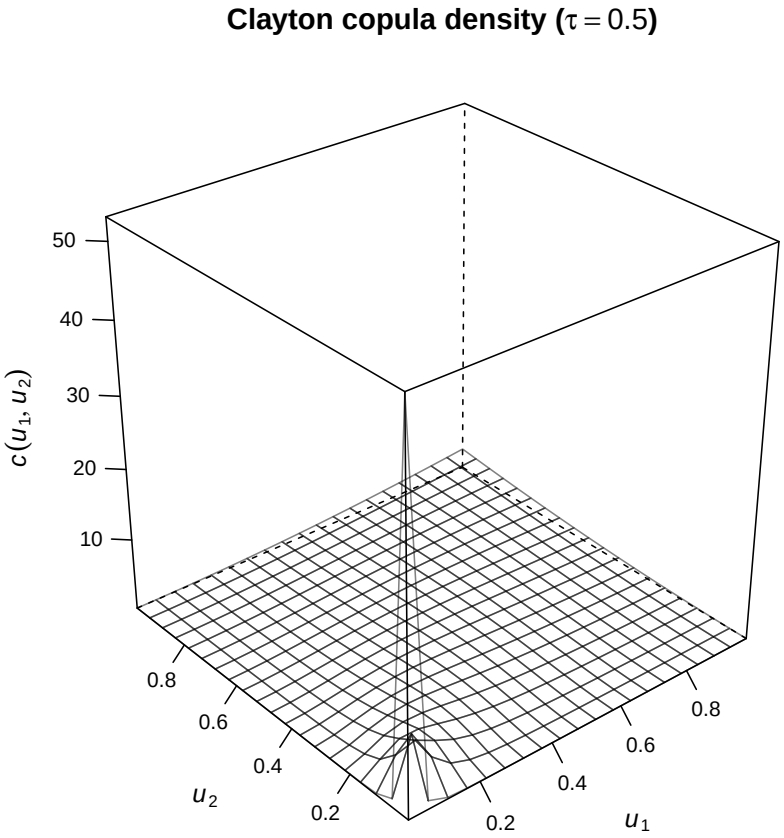
Right: Corresponding scatter plot (sample size $n = 1000$)

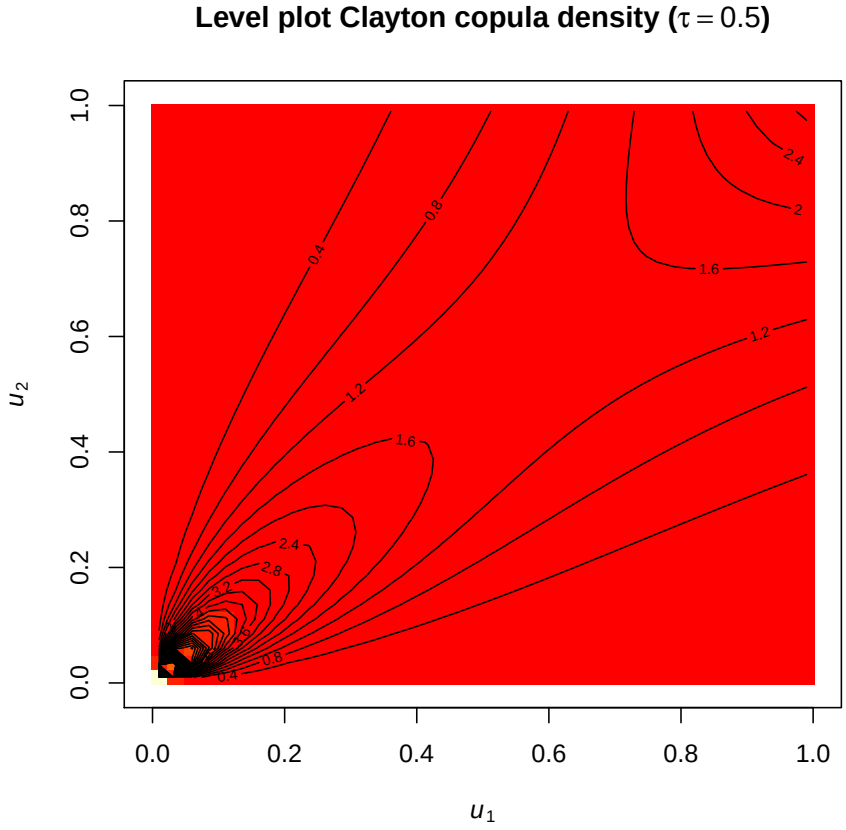




Left: Plot of the corresponding density.

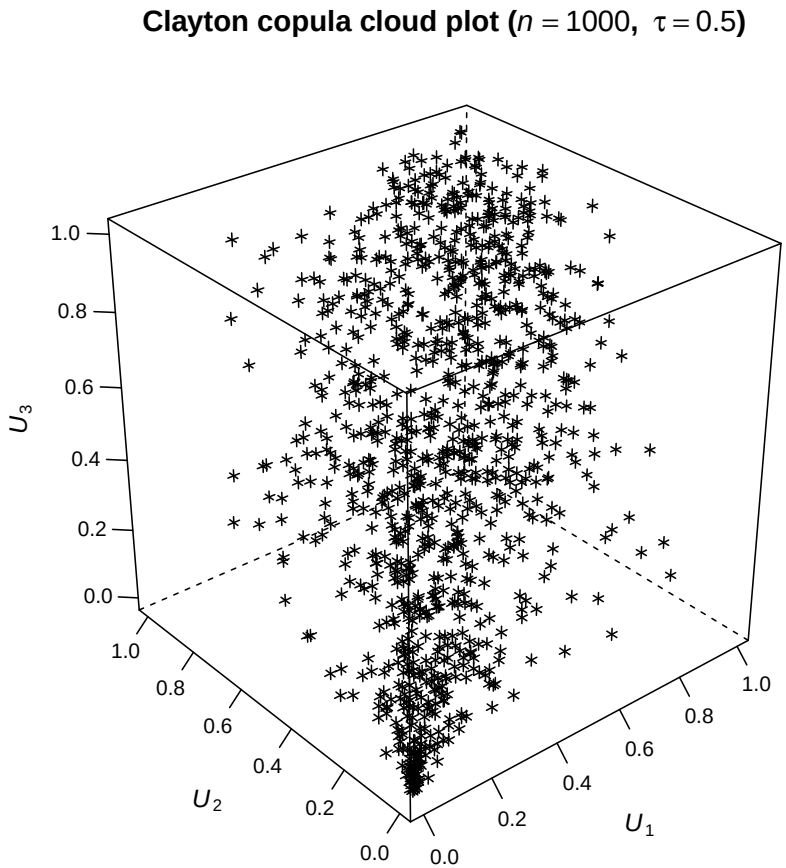
Right: Level plot of the density (with heat colors).

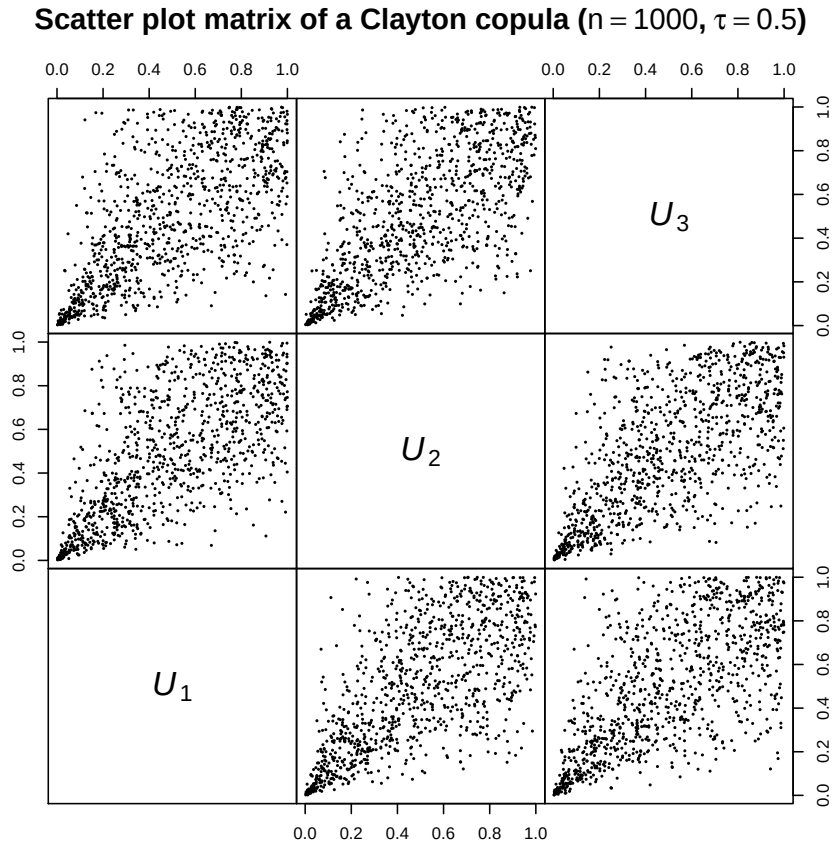




Left: Cloud plot of a trivariate Clayton copula (sample size $n = 1000$; Kendall's tau 0.5).

Right: Corresponding scatter plot matrix.

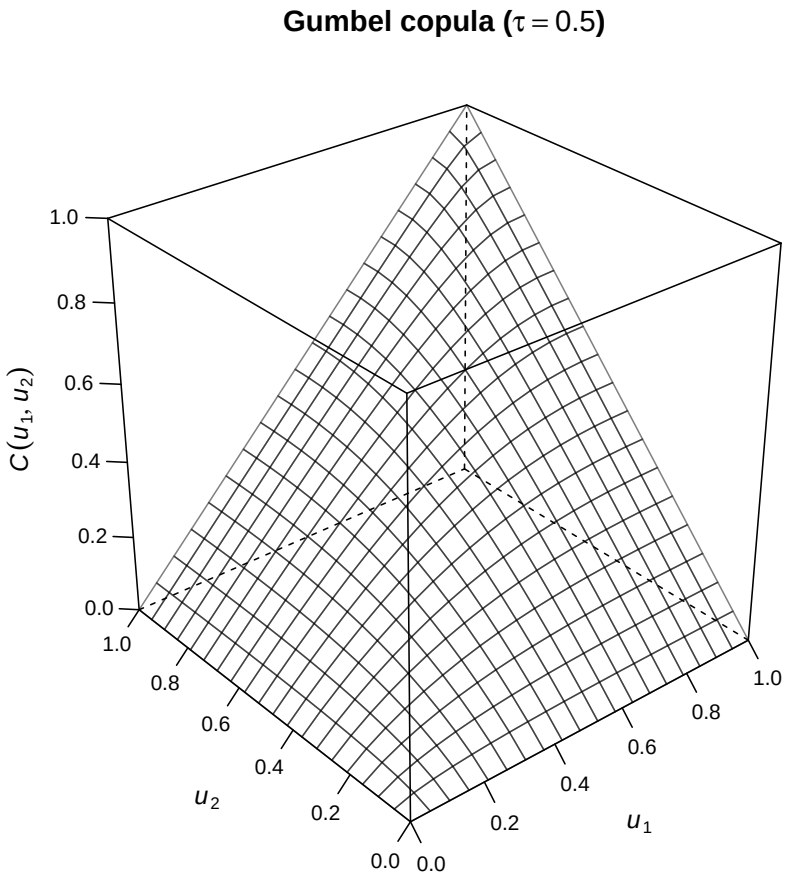


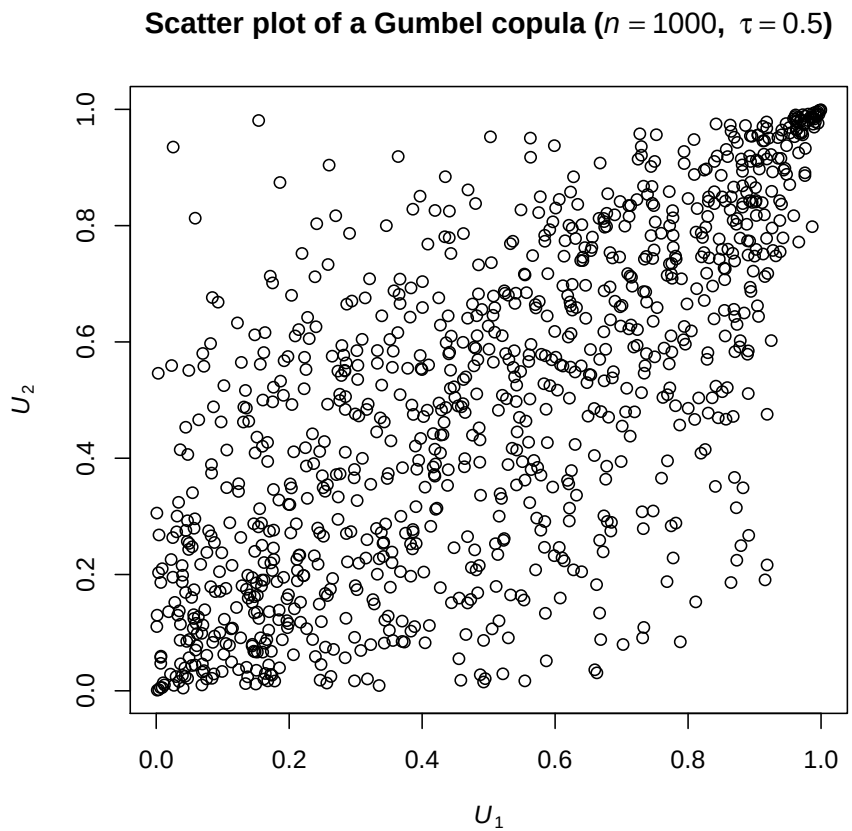


Gumbel copulas are obtained for $\psi(t) = \exp(-t^{1/\theta})$, $t \in [0, \infty)$, $\theta \in [1, \infty)$. For $\theta = 1$, $C = \Pi$; and for $\theta \to \infty$, $C \to M$.

Left: Plot of a bivariate Gumbel copula (Kendall's tau 0.5).

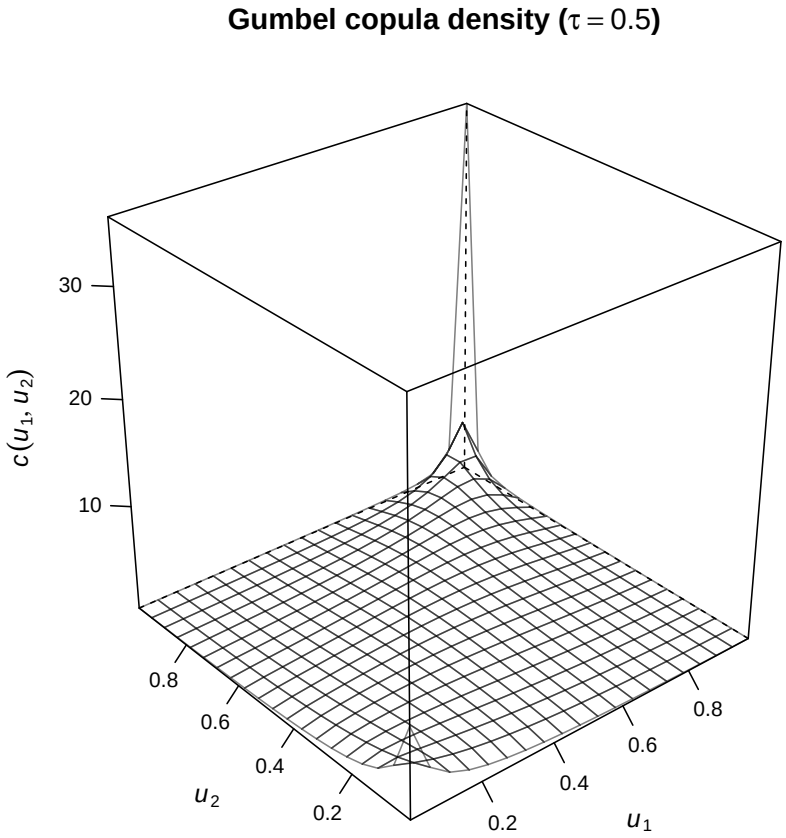
Right: Corresponding scatter plot (sample size $n = 1000$)

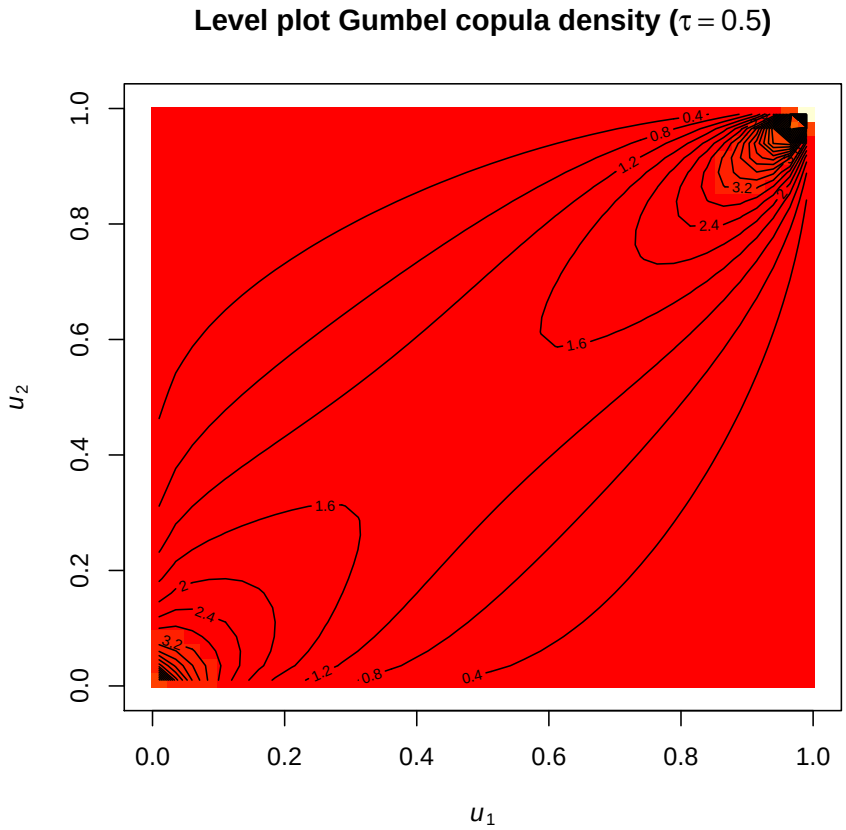




Left: Plot of the corresponding density.

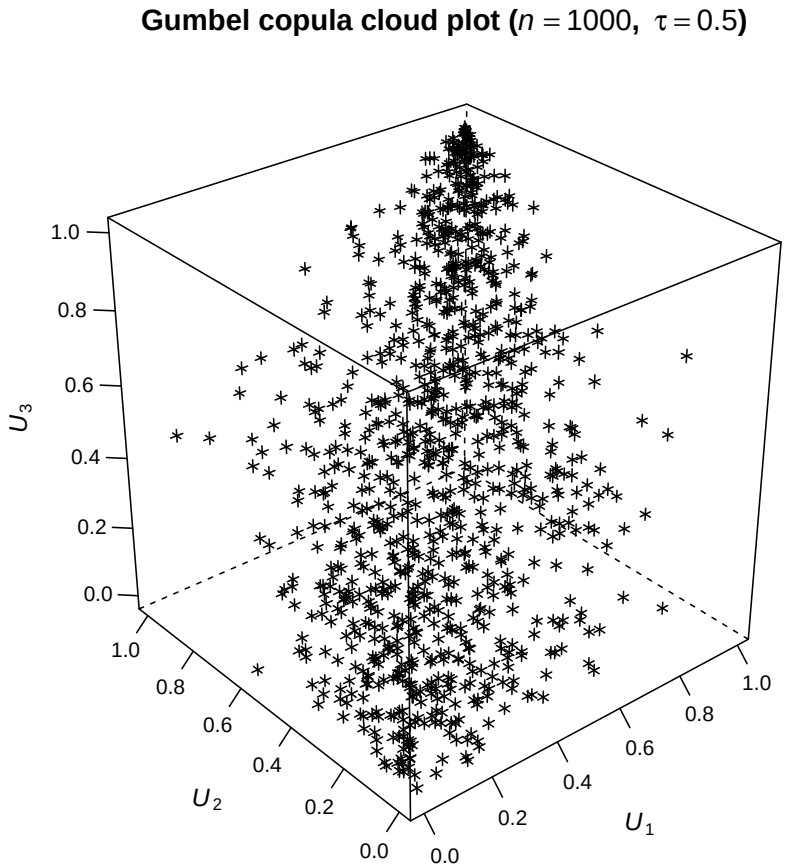
Right: Level plot of the density (with heat colors).

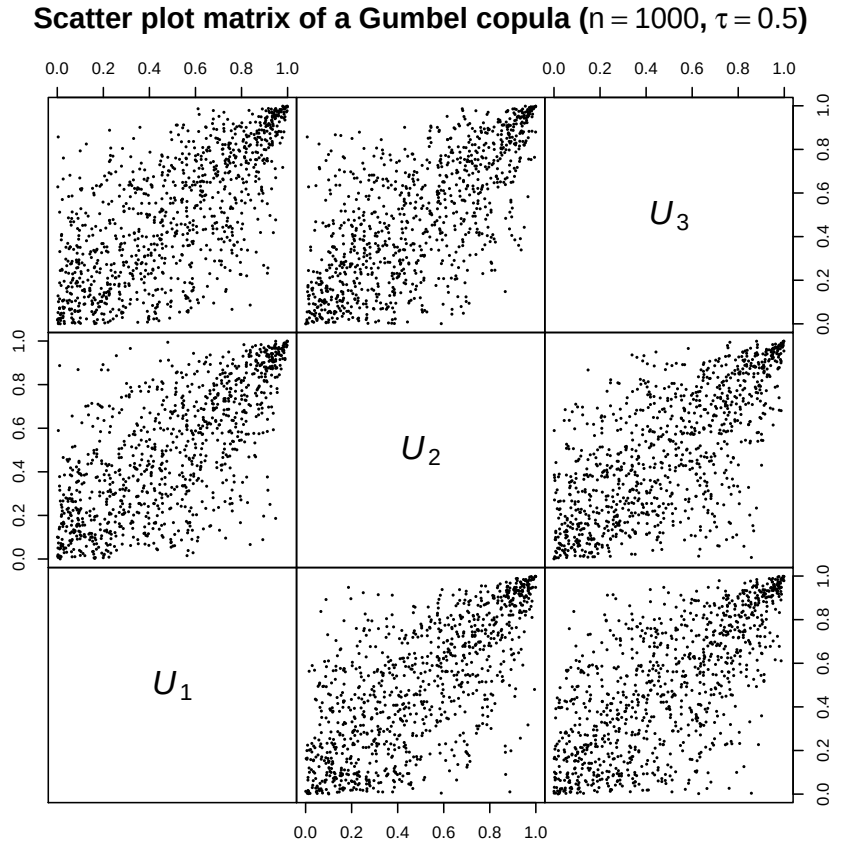




Left: Cloud plot of a trivariate Gumbel copula (sample size $n = 1000$; Kendall's tau 0.5).

Right: Corresponding scatter plot matrix.





Advantages and drawbacks of Archimedean copulas:

Advantages:

- Typically explicit (if ψ^{-1} is available)
- Useful in calculations: Properties can typically be expressed in terms of ψ
- Densities of various examples available
- Sampling often simple
- Not restricted to radial symmetry

Drawbacks:

- All margins of the same dimension are equal (symmetry or exchangeability)
- Often used only with a small number of parameters (some extensions available, but still less than $d(d-1)/2$)

7.1.3 Meta distributions

- *Fréchet class*: Class of all dfs F with given marginal dfs $F_1, \dots, F_d$;
  Meta-C models: All dfs F with the same given copula C.
- **Example:** A meta-t model is a multivariate df F with t copula C and some margins $F_1, \dots, F_d$.

7.1.4 Simulation of copulas and meta distributions

Copulas are typically sampled via specific stochastic representations.

Sampling implicit copulas

Algorithm 7.7 (Simulation of implicit copulas)

1) Sample $\boldsymbol{X} \sim F$, where F is a df with continuous margins $F_1, \dots, F_d$.
2) Return $\boldsymbol{U} = (F_1(X_1), \dots, F_d(X_d))$ (probability transformation).

Example 7.8

- Sampling Gauss copulas C_P^{Ga}:

  1) Sample $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{0}, P)$ ($\boldsymbol{X} \overset{\mathrm{d}}{=} A\boldsymbol{Z}$ for $AA' = P$, $\boldsymbol{Z} \sim \mathrm{N}_d(\boldsymbol{0}, I_d)$).

  2) Return $\boldsymbol{U} = (\Phi(X_1), \dots, \Phi(X_d))$.

- Sampling t_ν copulas $C_{\nu,P}^t$:

  1) Sample $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{0}, P)$ ($\boldsymbol{X} \overset{\mathrm{d}}{=} \sqrt{W} A\boldsymbol{Z}$ for $W = \frac{1}{V}$, $V \sim \Gamma(\frac{\nu}{2}, \frac{\nu}{2})$).

  2) Return $\boldsymbol{U} = (t_\nu(X_1), \dots, t_\nu(X_d))$.

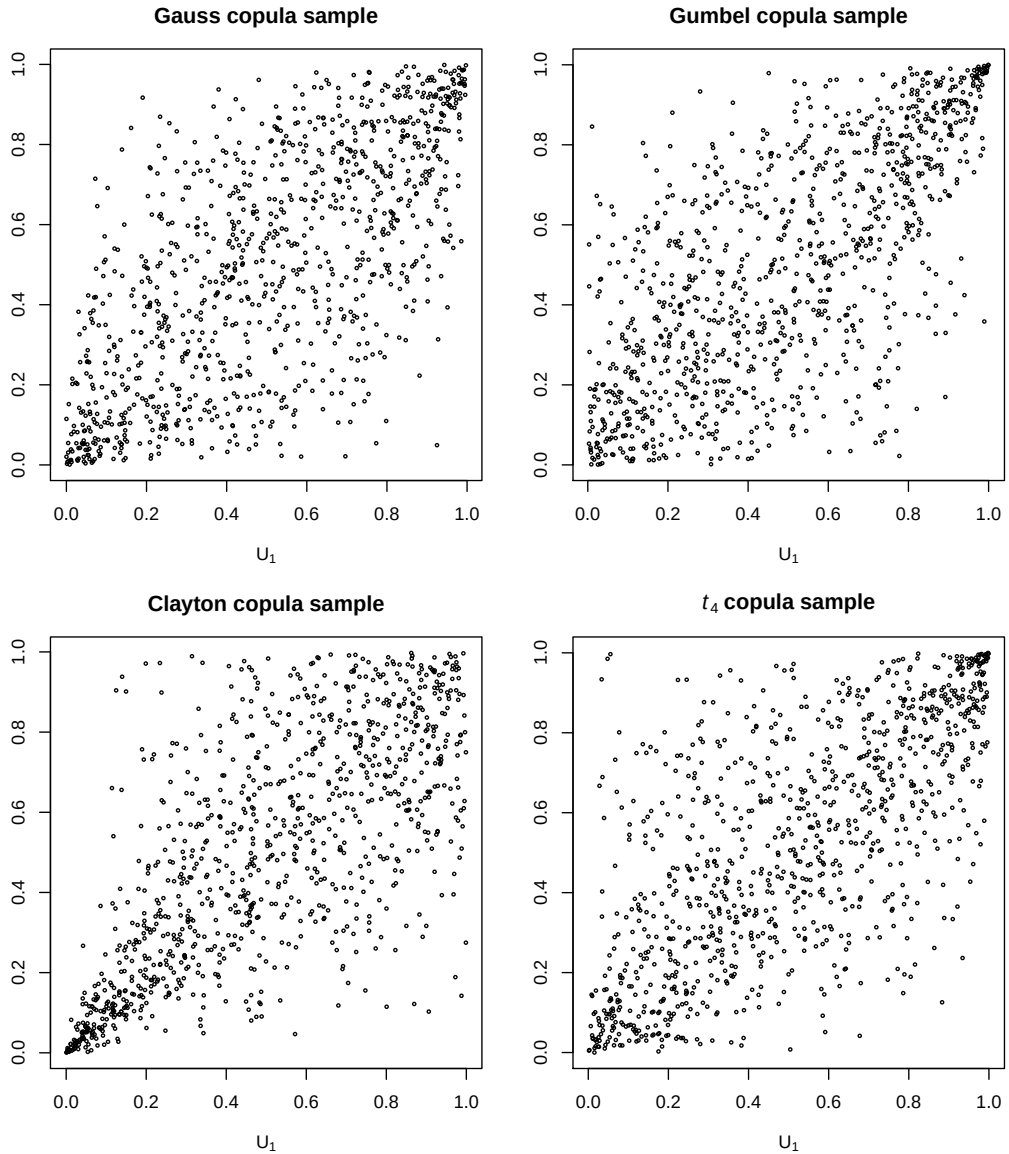
Sampling meta distributions

Meta-C distributions can be sampled via Sklar's Theorem, Part 2).

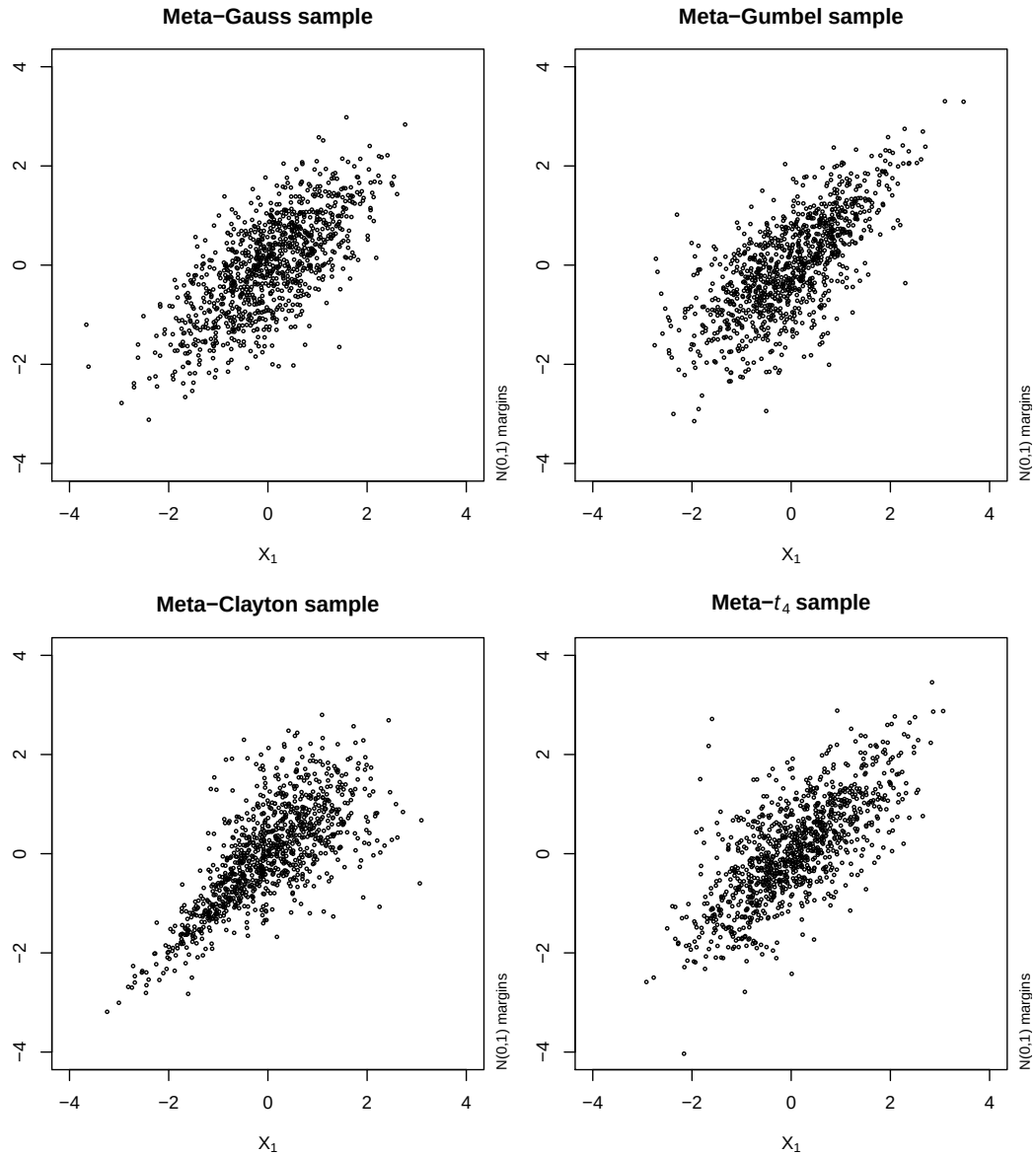
Algorithm 7.9 (Sampling meta-C models)

1) Sample $\boldsymbol{U} \sim C$.
2) Return $\boldsymbol{X} = (F_1^{\leftarrow}(U_1), \dots, F_d^{\leftarrow}(U_d))$ (quantile transformation).

2000 samples from (a): $C^{\mathsf{Ga}}_{\rho=0.7}$; (b): $C^{\mathsf{G}}_{\theta=2}$; (c): $C^{\mathsf{C}}_{\theta=2.2}$; (d): $C^{t}_{\nu=4,\,\rho=0.71}$



... transformed to $\mathrm{N}(0,1)$ margins; all have linear correlation ≈ 0.7!



7.1.5 Further properties of copulas

Copula densities

- By Sklar's Theorem, if F_j has density f_j, $j \in \{1, \dots d\}$, and C has density c, then the density f of F satisfies

$$f(\boldsymbol{x}) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{j=1}^{d} f_j(x_j). \tag{22}$$

  This implies

$$c(\boldsymbol{u}) = \frac{f(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdot \dots \cdot f_d(F_d^{-1}(u_d))}.$$

- It follows from (22) that the log-density splits into

$$\log f(\boldsymbol{x}) = \log c(F_1(x_1), \dots, F_d(x_d)) + \sum_{j=1}^{d} \log f_j(x_j)$$

  which allows for a *two-stage estimation* (marginal and copula parameters separately).

7.2 Dependence concepts and measures

Measures of association/dependence are scalar measures which summarize the dependence in terms of a single number. There are better and worse examples of such measures, which we will study in this section.

7.2.1 Perfect dependence

X_1, X_2 are *countermonotone* if (X_1, X_2) has copula W.

$X_1, \dots, X_d$ are *comonotone* if $(X_1, \dots, X_d)$ has copula M.

Proposition 7.10 (Perfect dependence)

1) $X_2 = T(X_1)$ a.s. with decreasing $T(x) = F_2^{\leftarrow}(1 - F_1(x))$ (countermonotone) if and only if $C(u_1, u_2) = W(u_1, u_2)$, $u_1, u_2 \in [0, 1]$.
2) $X_j = T_j(X_1)$ a.s. with increasing $T_j(x) = F_j^{\leftarrow}(F_1(x))$, $j \in \{2, \dots, d\}$, (comonotone) if and only if $C(\boldsymbol{u}) = M(\boldsymbol{u})$, $\boldsymbol{u} \in [0, 1]^d$.

7.2.2 Linear correlation

For two random variables X_1 and X_2 with $\mathbb{E}(X_j^2) < \infty$, $j \in \{1, 2\}$, the (*linear* or *Pearson's*) *correlation coefficient* ρ is defined by

$$\rho(X_1, X_2) = \frac{\operatorname{cov}(X_1, X_2)}{\sqrt{\operatorname{var} X_1}\sqrt{\operatorname{var} X_2}} = \frac{\mathbb{E}((X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2))}{\sqrt{\mathbb{E}((X_1 - \mathbb{E}X_1)^2)}\sqrt{\mathbb{E}((X_2 - \mathbb{E}X_2)^2)}}.$$

Classical properties and drawbacks of linear correlation

Let X_1 and X_2 be two random variables with $\mathbb{E}(X_j^2) < \infty$, $j \in \{1, 2\}$. Note that ρ depends on the marginal distributions! In particular, second moments have to exist (not the case, e.g. for $X_1, X_2 \overset{\text{ind.}}{\sim} F(x) = 1 - x^{-3/2}$!)

- $|\rho| \leq 1$. Furthermore, $|\rho| = 1$ if and only if there are constants $a \in \mathbb{R}\backslash\{0\}, b \in \mathbb{R}$ with $X_2 = aX_1 + b$ a.s. with $a \gtrless 0$ if and only if $\rho = \pm 1$. This discards other strong functional dependence such as $X_2 = X_1^2$, for example.

- If X_1 and X_2 are independent, then $\rho = 0$. However, the converse is not true in general; see Example 7.12 below.
- ρ is invariant under strictly increasing linear transformations on $\operatorname{ran} X_1 \times \operatorname{ran} X_2$ but not invariant under strictly increasing functions in general. To see this, consider $(X_1, X_2) \sim \mathrm{N}_2(\mathbf{0}, P)$. Then $\rho(X_1, X_2) = P_{12}$, but (as one can show) $\rho(F_1(X_1), F_2(X_2)) = \frac{6}{\pi} \arcsin(P_{12}/2)$.

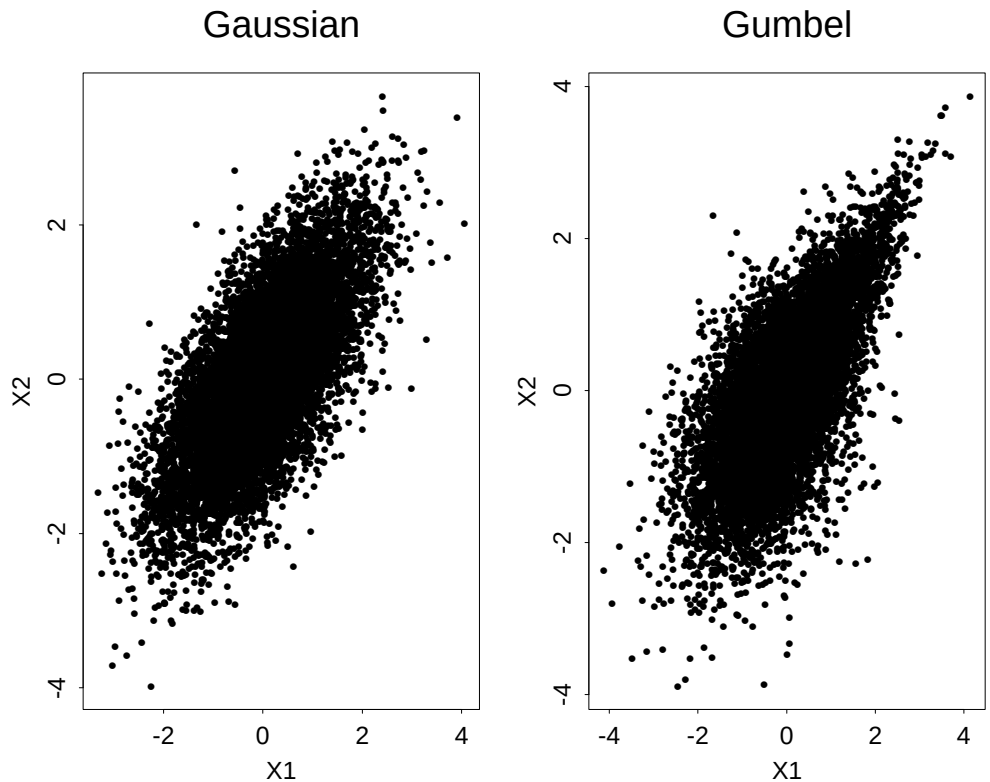
Proposition 7.11 (Hoeffding's formula)
Let $X_j \sim F_j$, $j \in \{1, 2\}$, be two random variables with $\mathbb{E}(X_j^2) < \infty$, $j \in \{1, 2\}$, and joint distribution function F. Then

$$\operatorname{cov}(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F(x_1, x_2) - F_1(x_1) F_2(x_2)) \, \mathrm{d}x_1 \mathrm{d}x_2.$$

Correlation fallacies

Fallacy 1: F_1, F_2, and ρ uniquely determine F

This is true for bivariate elliptical distributions, but wrong in general. The following samples both have $\mathrm{N}(0,1)$ margins and correlation $\rho = 0.7$, yet come from different (copula) models:



Another example is this.

Example 7.12 (Uncorrelated $\not\Rightarrow$ independent)

- Consider the two risks

$$X_1 = Z \quad \text{(Profit \& Loss Country A)},$$
$$X_2 = ZV \quad \text{(Profit \& Loss Country B)},$$

where V, Z are independent with $Z \sim \mathrm{N}(0,1)$ and $\mathbb{P}(V = -1) = \mathbb{P}(V = 1) = 1/2$. Then $X_2 \sim \mathrm{N}(0,1)$ and $\rho(X_1, X_2) = \mathrm{cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) \underset{\text{ind.}}{=} \mathbb{E}(V)\mathbb{E}(Z^2) = 0$, but X_1 and X_2 are not independent (in fact, V makes (X_1, X_2) switch between counter- and comonotonicity).

- Consider $(X_1', X_2') \sim \mathrm{N}_2(\mathbf{0}, I_2)$. Both (X_1', X_2') and (X_1, X_2) have $\mathrm{N}(0,1)$ margins and $\rho = 0$, but the copula of (X_1', X_2') is Π and the copula of (X_1, X_2) is the convex combination $C(\boldsymbol{u}) = \lambda M(\boldsymbol{u}) + (1 - \lambda)W(\boldsymbol{u})$ for $\lambda = 0.5$.

Fallacy 2: Given F_1, F_2, any $\rho \in [-1, 1]$ is attainable

This is true for elliptically distributed (X_1, X_2) with $\mathbb{E}(R^2) < \infty$ (as then $\operatorname{corr} \boldsymbol{X} = P$), but wrong in general:

- If F_1 and F_2 are not of the same type (no linearity), $\rho(X_1, X_2) = 1$ is not attainable (recall that $|\rho| = 1$ if and only if there are constants $a \in \mathbb{R}\backslash\{0\}, b \in \mathbb{R}$ with $X_2 = aX_1 + b$ a.s.).
- What is the attainable range then? Hoeffding's formula

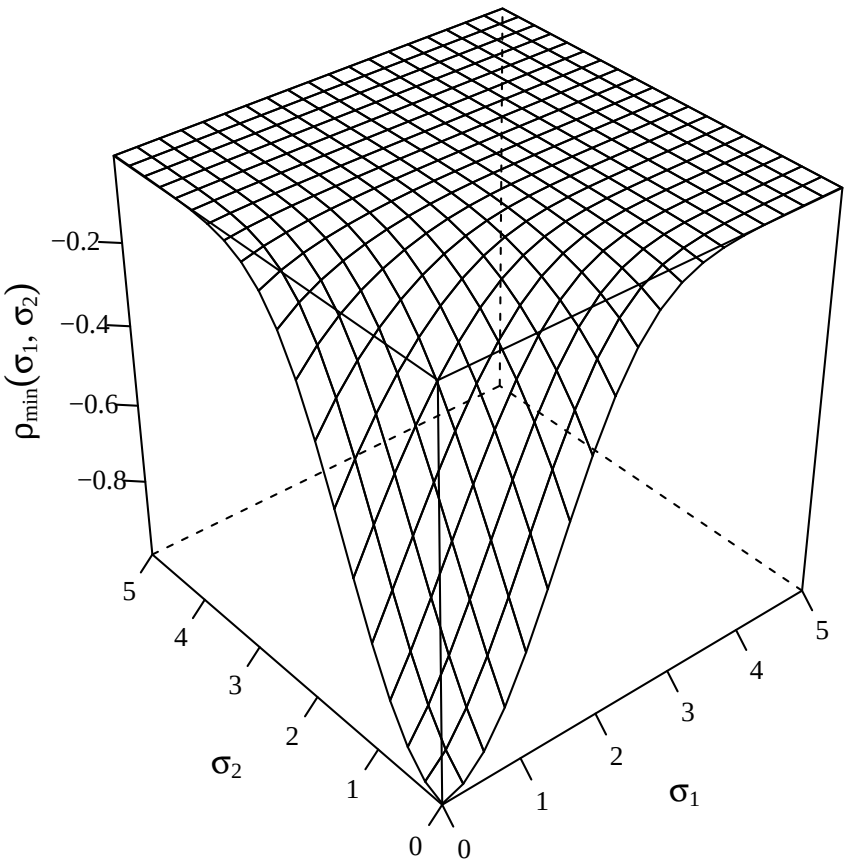
$$\operatorname{cov}(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (C(F_1(x_1), F_2(x_2)) - F_1(x_1)F_2(x_2)) \, \mathrm{d}x_1 \mathrm{d}x_2.$$

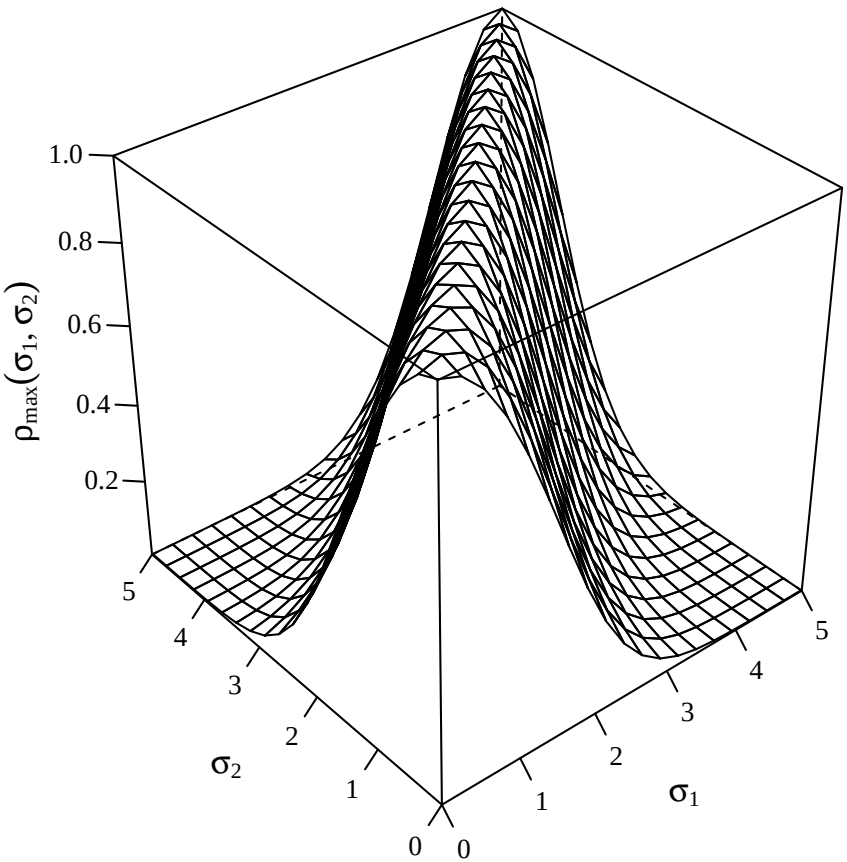
implies bounds on attainable ρ:

$\rho \in [\rho_{\min}, \ \rho_{\max}]$ ($\rho_{\min}$ is attained for $C = W$, $\rho_{\max}$ for $C = M$).

Example 7.13 (Bounds for a model with $\mathrm{LN}(0, \sigma_j^2)$ margins)

Let $X_j \sim \mathrm{LN}(0, \sigma_j^2)$, $j \in \{1, 2\}$. One can show that minimal (ρ_{min}; left) and maximal (ρ_{max}; right) correlations are given as follows.





For $\sigma_1^2 = 1$, $\sigma_2^2 = 16$ one has $\rho \in [-0.0003,\ 0.0137]$!

Fallacy 3: ρ maximal (i.e. $C = M$) $\Rightarrow \mathrm{VaR}_\alpha(X_1 + X_2)$ maximal

- This is true if (X_1, X_2) is elliptically distributed since the maximal $\rho = 1$ implies that X_1, X_2 are comonotone, so VaR_α can be shown to be additive and additivity provides the largest possible bound in this case as VaR_α is subadditive in this case.
- Any superadditivity example $\mathrm{VaR}_\alpha(X_1+X_2) > \mathrm{VaR}_\alpha(X_1)+\mathrm{VaR}_\alpha(X_2)$ serves as a counterexample as the right-hand side under comonotonicity (so maximal correlation) only *equals* $\mathrm{VaR}_\alpha(X_1 + X_2)$; see Section 2.3.5.

7.2.3 Rank correlation

Rank correlation coefficients are...

- ... always defined;
- ... invariant under strictly increasing transformations of the random variables (hence only depend on the underlying copula).

Kendall's tau and Spearman's rho

Definition 7.14 (Kendall's tau)
Let $X_j \sim F_j$ with F_j continuous, $j \in \{1, 2\}$. Let (X_1', X_2') be an independent copy of (X_1, X_2). *Kendall's tau* is defined by

$$\begin{aligned}\rho_\tau &= \mathbb{E}(\text{sign}((X_1 - X_1')(X_2 - X_2'))) \\ &= \mathbb{P}((X_1 - X_1')(X_2 - X_2') > 0) - \mathbb{P}((X_1 - X_1')(X_2 - X_2') < 0),\end{aligned}$$

where $\text{sign}(x) = I_{(0,\infty)}(x) - I_{(-\infty,0)}(x)$ (so -1 for $x < 0$, 0 for $x = 0$ and 1 for $x > 0$).

By definition, Kendall's tau is the probability of *concordance* ($\mathbb{P}((X_1 - X_1')(X_2 - X_2') > 0)$; probability of two independent points from F to have a positive slope) minus the probability of *discordance* ($\mathbb{P}((X_1 - X_1')(X_2 - X_2') < 0)$; probability of two independent points from F to have a negative slope).

Proposition 7.15 (Formula for Kendall's tau)

Let $X_j \sim F_j$ with F_j continuous, $j \in \{1, 2\}$, and copula C. Then

$$\rho_\tau = 4 \int_0^1 \int_0^1 C(u_1, u_2) \, \mathrm{d}C(u_1, u_2) - 1 = 4 \, \mathbb{E}(C(U_1, U_2)) - 1,$$

where $(U_1, U_2) \sim C$.

Definition 7.16 (Spearman's rho)

Let $X_j \sim F_j$ with F_j continuous, $j \in \{1, 2\}$. *Spearman's rho* is defined by $\rho_S = \rho(F_1(X_1), F_2(X_2))$.

Proposition 7.17 (Formula for Spearman's rho)

Let $X_j \sim F_j$ with F_j's continuous and copula C. For $(U_1', U_2') \sim \Pi$,

$$\rho_S = 12 \int_0^1 \int_0^1 C(u_1, u_2) \, \mathrm{d}u_1 \mathrm{d}u_2 - 3 = 12 \mathbb{E}(C(U_1', U_2')) - 3.$$

- Fallacy 1 (F_1, F_2, ρ uniquely determine F) is not solved by replacing ρ by rank correlation coefficients κ (it is easy to construct several copulas with the same Kendall's tau, e.g. via Archimedean copulas).
- Fallacy 2 (For F_1, F_2, any $\rho \in [-1, 1]$ is attainable) is solved when ρ is replaced by ρ_τ or ρ_S. Take

$$F(x_1, x_2) = \lambda M(F_1(x_1), F_2(x_2)) + (1 - \lambda) W(F_1(x_1), F_2(x_2)).$$

  This is a model with $\rho_\tau = \rho_S = 2\lambda - 1$ (choose $\lambda \in [0, 1]$ as desired).
- Fallacy 3 ($C = M$ implies $\mathrm{VaR}_\alpha(X_1 + X_2)$ maximal) is also not solved by rank correlation coefficients $\kappa = 1$: Although $\kappa = 1$ corresponds to $C = M$, this copula does not necessarily provide the largest $\mathrm{VaR}_\alpha(X_1 + X_2)$; see Fallacy 3 earlier.
- Nevertheless, rank correlations are useful to summarize dependence, to parameterize copula families to make dependence comparable and for copula parameter calibration or estimation.

7.2.4 Coefficients of tail dependence

Goal: Measure extremal dependence, i.e. dependence in the joint tails.

Definition 7.18 (Tail dependence)
Let $X_j \sim F_j$, $j \in \{1, 2\}$, be continuously distributed random variables. Provided that the limits exist, the *lower tail-dependence coefficient* λ_l and *upper tail-dependence coefficient* λ_u of X_1 and X_2 are defined by

$$\lambda_l = \lim_{u \downarrow 0} \mathbb{P}(X_2 \leq F_2^{\leftarrow}(u) \,|\, X_1 \leq F_1^{\leftarrow}(u)),$$

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

If $\lambda_l \in (0, 1]$ ($\lambda_u \in (0, 1]$), then (X_1, X_2) is *lower (upper) tail dependent*.
If $\lambda_l = 0$ ($\lambda_u = 0$), then (X_1, X_2) is *lower (upper) tail independent*.

As (conditional) probabilities, we clearly have $\lambda_l, \lambda_u \in [0, 1]$.

- Tail dependence is a copula property, since

$$\mathbb{P}(X_2 \le F_2^{\leftarrow}(u) \,|\, X_1 \le F_1^{\leftarrow}(u)) = \frac{\mathbb{P}(X_1 \le F_1^{\leftarrow}(u), X_2 \le F_2^{\leftarrow}(u))}{\mathbb{P}(X_1 \le F_1^{\leftarrow}(u))}$$

$$= \frac{F(F_1^{\leftarrow}(u), F_2^{\leftarrow}(u))}{F_1(F_1^{\leftarrow}(u))} \overset{\text{Sklar}}{=} \frac{C(u,u)}{u}, \ u \in (0,1), \text{ so } \lambda_l = \lim_{u \downarrow 0} \frac{C(u,u)}{u}.$$

- If $u \mapsto C(u,u)$ is differentiable in a neighborhood of 0 and the limit exists, then $\lambda_l = \lim_{u \downarrow 0} \frac{d}{du} C(u,u)$ (l'Hôpital's Rule).
- If C is totally differentiable in a neighborhood of 0 and the limit exists, then $\lambda_l = \lim_{u \downarrow 0} (\mathrm{D}_1 C(u,u) + \mathrm{D}_2 C(u,u))$ (Chain Rule). If C is exchangeable, $\lambda_l = 2 \lim_{u \downarrow 0} \mathrm{D}_1 C(u,u) = 2 \lim_{u \downarrow 0} C_{2|1}(u \,|\, u) = 2 \lim_{u \downarrow 0} \mathbb{P}(U_2 \le u \,|\, U_1 = u)$ for $(U_1, U_2) \sim C$. Combined with any continuous df F. (the same for both components) and $(X_1, X_2) = (F_\cdot^{\leftarrow}(U_1), F_\cdot^{\leftarrow}(U_2))$, one has

$$\lambda_l = 2 \lim_{x \downarrow -\infty} \mathbb{P}(X_2 \le x \,|\, X_1 = x). \tag{23}$$

  which is useful for deriving λ_l for elliptical copulas.

- Similarly as above, for the upper tail-dependence coefficient,

$$\lambda_u = \lim_{u\uparrow 1} \frac{1-2u+C(u,u)}{1-u}$$
$$= \lim_{u\uparrow 1} \frac{2(1-u)-(1-C(u,u))}{1-u} = 2 - \lim_{u\uparrow 1} \frac{1-C(u,u)}{1-u}.$$

- For all radially symmetric copulas (e.g. the bivariate C_P^{Ga} and $C_{\nu,P}^t$ copulas), we have $\lambda_l = \lambda_u =: \lambda$.

- For Archimedean copulas with strict ψ, a substitution and l'Hôpital's Rule show:

$$\lambda_l = \lim_{u\downarrow 0} \frac{\psi(2\psi^{-1}(u))}{u} = \lim_{t\to\infty} \frac{\psi(2t)}{\psi(t)} = 2\lim_{t\to\infty} \frac{\psi'(2t)}{\psi'(t)},$$

$$\lambda_u = 2 - \lim_{u\uparrow 1} \frac{1-\psi(2\psi^{-1}(u))}{1-u} = 2 - \lim_{t\downarrow 0} \frac{1-\psi(2t)}{1-\psi(t)} = 2 - 2\lim_{t\downarrow 0} \frac{\psi'(2t)}{\psi'(t)}.$$

Clayton: $\lambda_l = 2^{-1/\theta}$, $\lambda_u = 0$; Gumbel: $\lambda_l = 0$, $\lambda_u = 2 - 2^{1/\theta}$

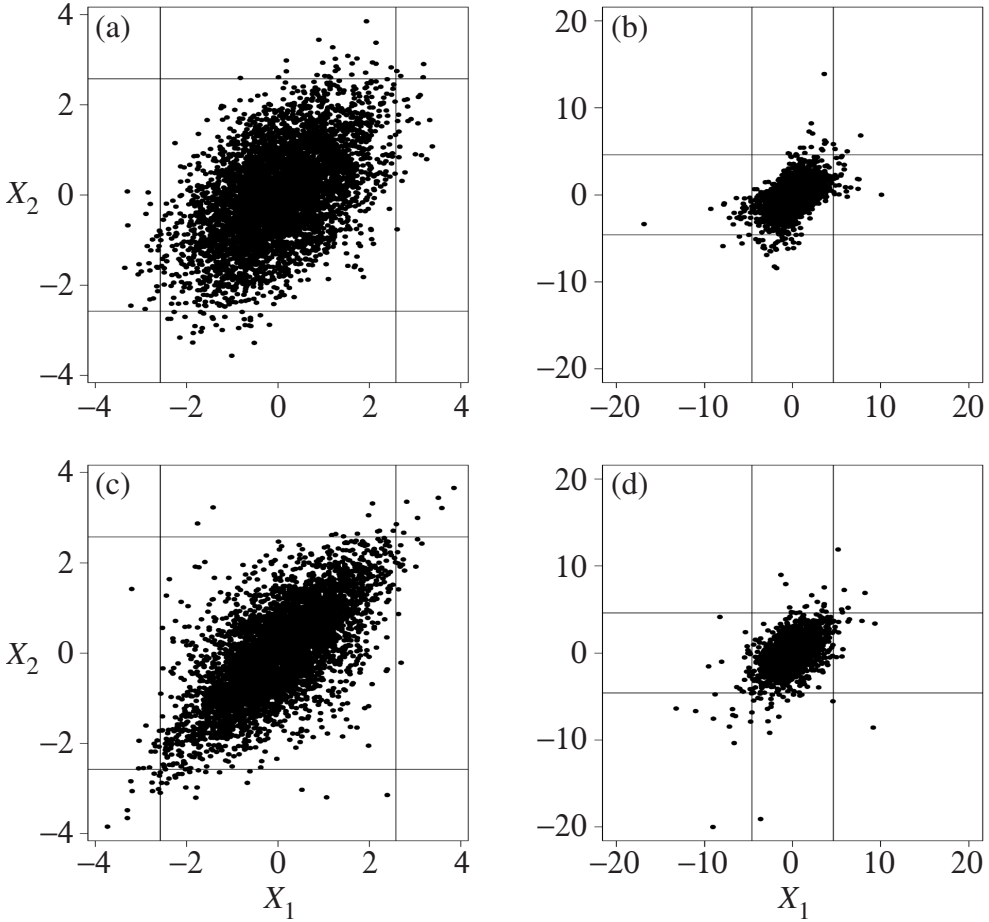
Example 7.19 (λ for the Gauss and t copula)

- Considering the bivariate $\mathrm{N}(\mathbf{0}, P)$ density, one can show (via $f_{X_2|X_1}(x_2 \,|\, x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)}$) that $(X_2 \,|\, X_1 = x) \sim \mathrm{N}(\rho x, 1-\rho^2)$. This implies that $\lambda = 2 \lim_{x \downarrow -\infty} \mathbb{P}(X_2 \leq x \,|\, X_1 = x) = 2 \lim_{x \downarrow -\infty} \Phi\left(\frac{x(1-\rho)}{\sqrt{1-\rho^2}}\right) = I_{\{\rho=1\}}$.
- For $C^t_{\nu,P}$, one can show that $(X_2 \,|\, X_1 = x) \sim t_{\nu+1}(\rho x, \frac{(1-\rho^2)(\nu+x^2)}{\nu+1})$ and thus $\mathbb{P}(X_2 \leq x \,|\, X_1 = x) = t_{\nu+1}\left(\frac{x-\rho x}{\sqrt{\frac{(1-\rho^2)(\nu+x^2)}{\nu+1}}}\right)$. Hence

  $\lambda = 2 t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right)$ (tail dependence; $\lambda \uparrow$ in $\rho \uparrow$ and $\nu \downarrow$).
- λ values for various ν, ρ:

ν	$\rho = -0.5$	$\rho = 0$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 1$
∞	0	0	0	0	1
10	0.00	0.01	0.08	0.46	1
4	0.01	0.08	0.25	0.63	1
2	0.06	0.18	0.39	0.72	1

Joint quantile exceedance probabilities



5000 samples from

(a) $\mathrm{N}_2(\mathbf{0}, P = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right))$, $\rho = 0.5$;

(b) C^{Ga}_ρ with t_4 margins (same dependence as in (a));

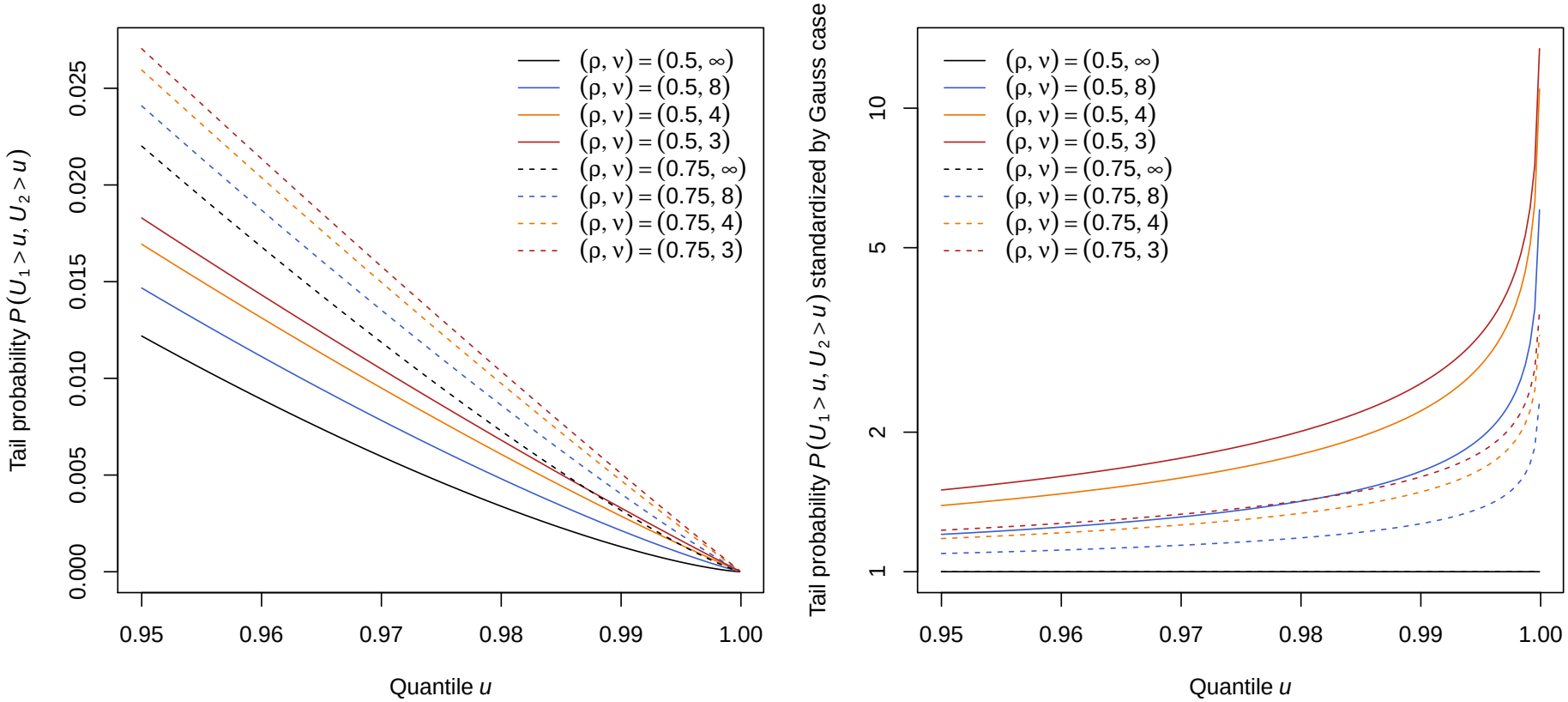
(c) $C^t_{4,\rho}$ with $\mathrm{N}(0,1)$ margins;

(d) $t_2(4, \mathbf{0}, P)$ (same dependence as in (c)).

Lines denote the true marginal 0.005- and 0.995-quantiles.

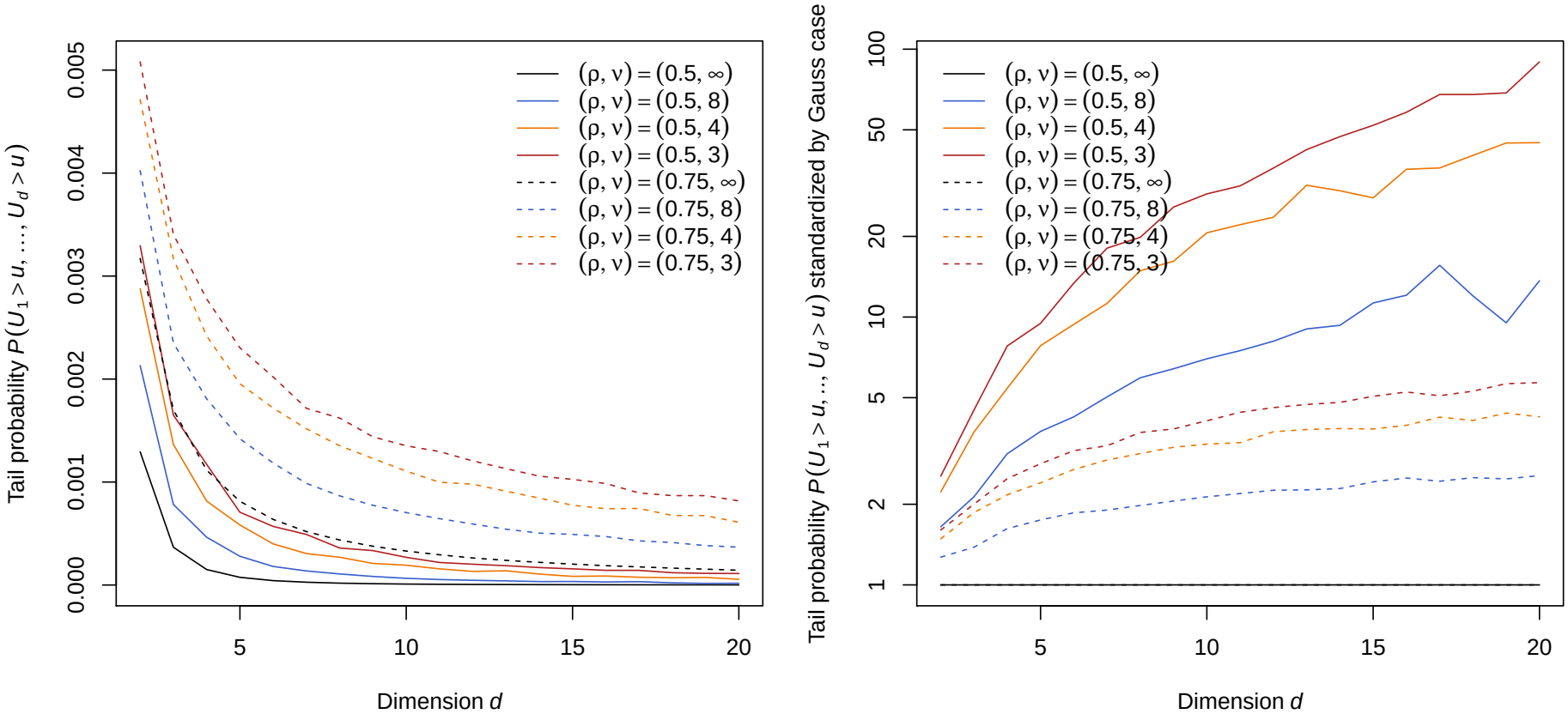
Note the different number of points in the bivariate tails (all models have the same Kendall's tau!)

Joint tail probabilities $\mathbb{P}(U_1 > u, U_2 > u)$ for $d = 2$



- **Left:** The higher ρ or the smaller ν, the larger $\mathbb{P}(U_1 > u, U_2 > u)$.
- **Right:** $u \mapsto \dfrac{\mathbb{P}(U_1 > u, U_2 > u)}{\mathbb{P}(V_1 > u, V_2 > u)} \underset{\text{symm.}}{\overset{\text{radial}}{=}} \dfrac{C^t_{\nu,\rho}(u,u)}{C^{\text{Ga}}_{\rho}(u,u)}$

Joint tail probabilities $\mathbb{P}(U_1 > u, \dots, U_d > u)$ for $u = 0.99$



- Homogeneous P (off-diagonal entry ρ). Note the MC randomness.
- **Left:** Clear; less mass in corners in higher dimensions.
- **Right:** $d \mapsto \dfrac{\mathbb{P}(U_1 > u, \dots, U_d > u)}{\mathbb{P}(V_1 > u, \dots, V_d > u)} \underset{\text{symm.}}{\overset{\text{radial}}{=}} \dfrac{C^t_{\nu,\rho}(u, \dots, u)}{C^{\text{Ga}}_{\rho}(u, \dots, u)}$ for $u = 0.99$.

Example 7.20 (Interpretation of joint tail probabilities)

- Consider 5 daily negative log-returns $\boldsymbol{X} = (X_1, \dots, X_5)$. Assume they follow an elliptical distribution and have pairwise correlations $\rho = 0.5$. However, we are unsure about the best joint model.
- If $\boldsymbol{X}$ are multivariate normal (and thus $C^{\text{Ga}}_{\rho=0.5}$), the probability that on any day all 5 negative returns lie above their $u = 0.99$ quantiles is

$$\mathbb{P}(X_1 > F_1^{\leftarrow}(u), \dots, X_5 > F_5^{\leftarrow}(u)) = \mathbb{P}(U_1 > u, \dots, U_5 > u) \underset{\text{MC error}}{\approx} 7.48 \times 10^{-5}.$$

  In the long run such an event will happen once every $1/7.48 \times 10^{-5} \approx 13\,369$ trading days on average ($\approx$ once every 51.4 years; assuming 260 trading days in a year).
- If $\boldsymbol{X}$ is multivariate t_3 (and thus $C^{t}_{\nu=3,\rho=0.5}$), however, such an event will happen approximately 10 times more often, i.e. $\approx$ once every 5.14 years. This gets worse the larger d!

7.2.5 Copulas and credit risk

Felix Salmon: "Recipe for Disaster: The Formula That Killed Wall Street"

Here's what killed your 401(k) *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick—and fatally flawed—way to assess risk. A shorter version appears on this month's cover of* Wired.

Probability
Specifically, this is a joint default probability—the likelihood that any two members of the pool (A and B) will both default. It's what investors are looking for, and the rest of the formula provides the answer.

Survival times
The amount of time between now and when A and B can be expected to default. Li took the idea from a concept in actuarial science that charts what happens to someone's life expectancy when their spouse dies.

Equality
A dangerously precise concept, since it leaves no room for error. Clean equations help both quants and their managers forget that the real world contains a surprising amount of uncertainty, fuzziness, and precariousness.

Copula
This couples (hence the Latinate term copula) the individual probabilities associated with A and B to come up with a single number. Errors here massively increase the risk of the whole equation blowing up.

Distribution functions
The probabilities of how long A and B are likely to survive. Since these are not certainties, they can be dangerous: Small miscalculations may leave you facing much more risk than the formula indicates.

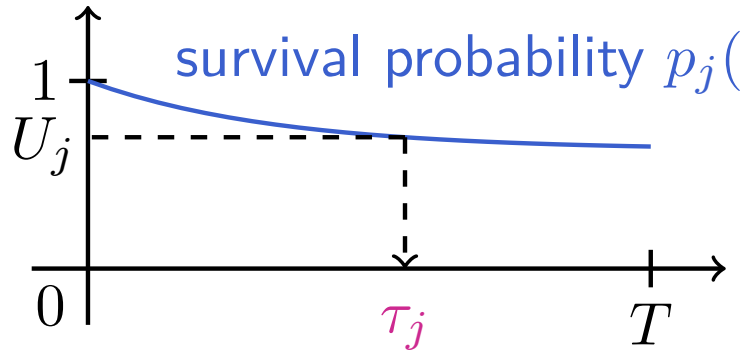
Gamma
The all-powerful correlation parameter, which reduces correlation to a single constant—something that should be highly improbable, if not impossible. This is the magic number that made Li's copula function irresistible.

How intensity-/copula-based default models work

Intensity-based default model:

$$p_j(t) = \exp\Big(-\int_0^t \lambda_j(s)\,ds\Big)$$

$$\tau_j = \inf\{t \geq 0 : p_j(t) \leq U_j\}$$



Note: $\lambda_U = 0$ (as for the Gauss copula!)

$\Rightarrow$ (Almost) no joint defaults! (p_j typically very flat)

Copulas for the triggers $\boldsymbol{U}$:

1) Li (2000): Gauss (Sibuya (1960): $\lambda_U = 0$)

2) Schönbucher and Schubert (2001): Archimedean ($\lambda_U > 0$)

3) Hofert and Scherer (2011): nested Archimedean ($\lambda_U > 0$, hierarchies)

Typical application: CDO pricing models based on iTraxx data.

7.3 A proof for subadditivity of ES

Proposition 7.21 (Subadditivity of ES)
$\mathrm{ES}_\alpha(L) = \frac{\sup_{\{\tilde{Y}\sim \mathrm{B}(1,1-\alpha)\}} \mathbb{E}(L\tilde{Y})}{1-\alpha}$, which is subadditive; the supremum is taken over all copulas between $L \sim F_L$ and $\tilde{Y} \sim \mathrm{B}(1, 1-\alpha)$.

Proof.

- Let $L = F_L^{\leftarrow}(U)$ and $Y = I_{\{U>\alpha\}} \sim \mathrm{B}(1, 1-\alpha)$ for $U \sim \mathrm{U}(0,1)$.
- Then $\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 F_L^{\leftarrow}(u)\,\mathrm{d}u = \frac{1}{1-\alpha}\int_0^1 F_L^{\leftarrow}(u) I_{\{u>\alpha\}} \cdot 1\,\mathrm{d}u = \frac{1}{1-\alpha}\mathbb{E}(F_L^{\leftarrow}(U) I_{\{U>\alpha\}}) = \frac{1}{1-\alpha}\mathbb{E}(LY)$.
- L and Y are comontone. For any other $(L, \tilde{Y})$ with $\tilde{Y} \sim \mathrm{B}(1, 1-\alpha)$,

$$\mathbb{E}(L\tilde{Y}) = \mathrm{cov}(L, \tilde{Y}) + \mathbb{E}(L)\mathbb{E}(\tilde{Y}) \underset{\text{Hoeffding}}{\leq} \mathrm{cov}(L, Y) + \mathbb{E}(L)\mathbb{E}(Y) = \mathbb{E}(LY)$$

and thus $\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha} \sup_{\{\tilde{Y}\sim \mathrm{B}(1,1-\alpha)\}} \mathbb{E}(L\tilde{Y})$. □

7.4 Fitting copulas to data

- Let $\boldsymbol{X}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{ind.}}{\sim} F$ with cont. margins $F_1, \ldots, F_d$ and copula C.
- We assume that we have data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, interpreted as realizations of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$; in what follows we work with the latter.
- Assume
  - $F_j = F_j(\cdot; \boldsymbol{\theta}_{0,j})$ for some $\boldsymbol{\theta}_{0,j} \in \Theta_j$, $j \in \{1, \ldots, d\}$;
    ($F_j(\cdot; \boldsymbol{\theta}_j)$ is assumed to be continuous $\forall\, \boldsymbol{\theta}_j \in \Theta_j$, $j \in \{1, \ldots, d\}$)
  - $C = C(\cdot; \boldsymbol{\theta}_{0,C})$ for some $\boldsymbol{\theta}_{0,C} \in \Theta_C$.

  Thus F has the true but unknown parameter vector $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{0,C}', \boldsymbol{\theta}_{0,1}', \ldots, \boldsymbol{\theta}_{0,d}')'$ to be estimated.
- Here, we focus particularly on $\boldsymbol{\theta}_{0,C}$. Whenever necessary, we assume that the margins $F_1, \ldots, F_d$ and the copula C are absolutely continuous with corresponding densities $f_1, \ldots, f_d$ and c, respectively.
- We assume the chosen copula to be appropriate (w.r.t. symmetry etc.).

7.4.1 Method-of-moments using rank correlation

- For $d = 2$ and one-parameter copulas, Genest and Rivest (1993) suggested estimating $\theta_{0,C}$ by solving $\rho_\tau(\theta_C) = r_n^\tau$ w.r.t. θ_C, i.e.

$$\hat{\theta}_{n,C}^{\text{IKTE}} = \rho_\tau^{-1}(r_n^\tau), \quad \text{(inversion of Kendall's tau estimator (IKTE))}$$

  where $\rho_\tau(\cdot)$ denotes Kendall's tau as a function of θ and r_n^τ is the sample version of Kendall's tau (computed from $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ or pseudo-observations $\boldsymbol{U}_1, \dots, \boldsymbol{U}_n$; see later).

- The standardized dispersion matrix P for elliptical copulas can be estimated via *pairwise inversion of Kendall's tau*. If $r_{n,j_1j_2}^\tau$ denotes the sample version of Kendall's tau for data pair (j_1, j_2), then

$$\hat{P}_{n,j_1j_2}^{\text{IKTE}} = \sin\left(\frac{\pi}{2} r_{n,j_1j_2}^\tau\right).$$

  A proper correlation matrix P can be constructed as in Higham (2002).

- One can also use Spearman's rho. For Gauss copulas,

$$\rho \approx \frac{6}{\pi} \arcsin \frac{\rho}{2} = \rho_S.$$

The approximation error is comparably small, so that the matrix of pairwise sample versions of Spearman's rho is an estimator for P.

- For t copulas, $\hat{P}_n^{\text{IKTE}}$ can be used to estimate P and then ν can be estimated via its MLE based on $\hat{P}_n^{\text{IKTE}}$; see Mashal and Zeevi (2002).

7.4.2 Forming a pseudo-sample from the copula

- $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ typically does not have $\mathrm{U}(0,1)$ margins. For applying the "copula approach" we thus need *pseudo-observations* from C.
- In general, we take $\hat{\boldsymbol{U}}_i = (\hat{U}_{i1}, \dots, \hat{U}_{id}) = (\hat{F}_1(X_{i1}), \dots, \hat{F}_d(X_{id}))$, $i \in \{1, \dots, n\}$, where $\hat{F}_j$ denotes an estimator of F_j. Note that $\hat{\boldsymbol{U}}_1, \dots, \hat{\boldsymbol{U}}_n$ are typically neither independent (even if $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ are) nor perfectly $\mathrm{U}(0,1)^d$ distributed.
- Possible choices for $\hat{F}_j$:

- Parametric estimators (typically if n is small). One often still uses (24) below for estimating $\boldsymbol{\theta}_{0,C}$ (to keep the error due to misspecification of the margins small).
- Semi-parametric estimators (for example EVT-based: Bodies are modelled empirically, tails semiparametrically via the GPD-based tail estimator of Smith (1987)).
- Non-parametric estimators with scaled empirical dfs, so

$$\hat{U}_{ij} = \frac{n}{n+1}\hat{F}_{n,j}(X_{ij}) = \frac{R_{ij}}{n+1}, \tag{24}$$

where R_{ij} denotes the *rank* of X_{ij} among all $X_{1j}, \dots, X_{nj}$. The scaling is to avoid density evaluation on the boundary of $[0,1]^d$.

If n is sufficiently large, one typically uses (24).

7.4.3 Maximum likelihood estimation

The (classical) maximum likelihood estimator

- If it exists, the density of $F(\boldsymbol{x}) = C(F_1(x_1), \dots, F_d(x_d))$ is

$$f(\boldsymbol{x}; \boldsymbol{\theta}_0) = c(F_1(x_1; \boldsymbol{\theta}_{0,1}), \dots, F_d(x_d; \boldsymbol{\theta}_{0,d}); \boldsymbol{\theta}_{0,C}) \prod_{j=1}^{d} f_j(x_j; \boldsymbol{\theta}_{0,j}).$$

- The log-likelihood based on $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ is thus

$$\ell(\boldsymbol{\theta}; \boldsymbol{X}_1, \dots, \boldsymbol{X}_n) = \sum_{i=1}^{n} \ell(\boldsymbol{\theta}; \boldsymbol{X}_i)$$

$$= \sum_{i=1}^{n} \ell_C(\boldsymbol{\theta}_C; F_1(X_{i1}; \boldsymbol{\theta}_1), \dots, F_d(X_{id}; \boldsymbol{\theta}_d)) + \sum_{i=1}^{n} \sum_{j=1}^{d} \ell_j(\boldsymbol{\theta}_j; X_{ij}),$$

where

$$\ell_C(\boldsymbol{\theta}_C; u_1, \dots, u_d) = \log c(u_1, \dots, u_d; \boldsymbol{\theta}_C)$$

$$\ell_j(\boldsymbol{\theta}_j; x) = \log f_j(x; \boldsymbol{\theta}_j), \quad j \in \{1, \dots, d\}.$$

- The *maximum likelihood estimator (MLE)* of $\boldsymbol{\theta}_0$ is

$$\hat{\boldsymbol{\theta}}_n^{\text{MLE}} = \operatorname*{argsup}_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}; \boldsymbol{X}_1, \dots, \boldsymbol{X}_n).$$

  This optimization is typically done by numerical means. Note that this can be quite demanding, especially in high dimensions.

The inference functions for margins estimator

- Joe and Xu (1996) suggested the two-step estimation approach:

  Step 1: For $j \in \{1, \dots, d\}$, estimate $\boldsymbol{\theta}_{0,j}$ by its MLE $\hat{\boldsymbol{\theta}}_{n,j}^{\text{MLE}}$.

  Step 2: Estimate $\boldsymbol{\theta}_{0,C}$ by

$$\hat{\boldsymbol{\theta}}_{n,C}^{\text{IFME}} = \operatorname*{argsup}_{\boldsymbol{\theta}_C\in\Theta_C} \ell(\boldsymbol{\theta}_C, \hat{\boldsymbol{\theta}}_{n,1}^{\text{MLE}}, \dots, \hat{\boldsymbol{\theta}}_{n,d}^{\text{MLE}}; \boldsymbol{X}_1, \dots, \boldsymbol{X}_n).$$

  The *inference functions for margins estimator (IFME)* of $\boldsymbol{\theta}_0$ is thus

$$\hat{\boldsymbol{\theta}}_n^{\text{IFME}} = (\hat{\boldsymbol{\theta}}_{n,C}^{\text{IFME}}, \hat{\boldsymbol{\theta}}_{n,1}^{\text{MLE}}, \dots, \hat{\boldsymbol{\theta}}_{n,d}^{\text{MLE}})$$

- This is typically much easier to compute than $\hat{\boldsymbol{\theta}}_n^{\text{MLE}}$.

Example 7.22 (A computationally convincing example)

Suppose $X_j \sim \mathrm{N}(\mu_j, \sigma_j^2)$, $j \in \{1, \dots, d\}$, for $d = 100$, and C has (just) one parameter.

1) MLE requires to solve a 201-dimensional optimization problem.
2) IFME only requires 100 optimizations in two dimensions and 1 one-dimensional optimization.

- If the marginals are estimated parametrically one often still uses the pseudo-observations built from the marginal empirical dfs to estimate $\boldsymbol{\theta}_{0,C}$ (see MPLE below) in order to avoid misspecifiation of the margins.
- In this case (and under more complicated marginal models), one can execute the 101 optimizations in parallel, independently of each other.

The maximum pseudo-likelihood estimator

- The *maximum pseudo-likelihood estimator (MPLE)*, introduced by Genest, Ghoudi, et al. (1995), works similarly to $\hat{\boldsymbol{\theta}}_n^{\text{IFME}}$, but estimates the margins non-parametrically:

  Step 1: Compute rank-based pseudo-observations $\hat{\boldsymbol{U}}_1, \dots, \hat{\boldsymbol{U}}_n$.

  Step 2: Estimate $\boldsymbol{\theta}_{0,C}$ by

$$\hat{\boldsymbol{\theta}}_{n,C}^{\text{MPLE}} = \operatorname*{argsup}_{\boldsymbol{\theta}_C \in \Theta_C} \sum_{i=1}^{n} \ell_C(\boldsymbol{\theta}_C; \hat{U}_{i1}, \dots, \hat{U}_{id}) = \operatorname*{argsup}_{\boldsymbol{\theta}_C \in \Theta_C} \sum_{i=1}^{n} \log c(\hat{\boldsymbol{U}}_i; \boldsymbol{\theta}_C).$$

- Kim et al. (2007) compare $\hat{\boldsymbol{\theta}}_n^{\text{MLE}}$, $\hat{\boldsymbol{\theta}}_n^{\text{IFME}}$, and $\hat{\boldsymbol{\theta}}_{n,C}^{\text{MPLE}}$ in a simulation study ($d = 2$ only!) and argue in favor of $\hat{\boldsymbol{\theta}}_{n,C}^{\text{MPLE}}$ overall, especially w.r.t. robustness against misspecification of the margins; but see Embrechts and Hofert (2013) for $d \gg 2$.

Example 7.23 (Fitting the Gauss copula)

- Use pairwise inversion of Spearman's rho or Kendall's tau.
- Or the MPLE via the (copula-related) log-likelihood

$$\ell_C(P; \hat{\boldsymbol{U}}_1, \dots, \hat{\boldsymbol{U}}_n) = \sum_{i=1}^n \ell_C(P; \hat{\boldsymbol{U}}_i) \underset{\text{Eq. (21)}}{=} \sum_{i=1}^n \log c_P^{\mathsf{Ga}}(\hat{\boldsymbol{U}}_i).$$

For maximization over all correlation matrices P, we can use the Cholesky factor A as reparameterization and maximize over all lower triangular matrices A with 1s on the diagonal.

Example 7.24 (Fitting the t copula)

- For small d, maximize the likelihood over all correlation matrices (as for the Gauss copula case) and the d.o.f. ν.
- For moderate/larger d, use Mashal and Zeevi (2002):

  1) Estimate P via pairwise inversion of Kendall's tau (see above).

  2) Plug $\hat{P}$ into the likelihood and maximize it w.r.t. ν to obtain $\hat{\nu}_n$.

Estimation is only one side of the coin. The other is *goodness-of-fit* (i.e. to find out whether our estimated model indeed represents the given data well) and model selection (i.e. to decide which model is best among all adequate fitted models). Goodness-of-fit can be (computationally) challenging, particularly for large d. There are also graphical approaches not further discussed here.

8 Aggregate risk

8.1 Coherent and convex risk measures

8.2 Law-invariant coherent risk measures

8.3 Risk measures for linear portfolios

8.4 Risk aggregation

8.5 Capital allocation

8.1 Coherent and convex risk measures

- Consider a linear space $\mathcal{M} \subseteq \mathcal{L}^0(\Omega, \mathcal{F}, \mathbb{P})$ (a.s. finite rvs).
- Each $L \in \mathcal{M}$ (incl. constants) represents a loss over a fixed time horizon.
- A *risk measure* is a mapping $\varrho : \mathcal{M} \to \mathbb{R}$; $\varrho(L)$ gives the total amount of capital needed to back a position with loss L.
- $C \subseteq \mathcal{M}$ is *convex* if $(1-\gamma)x + \gamma y \in C$ for all $x, y \in C$, $0 < \gamma < 1$. C is a *convex cone* if, additionally, $\lambda x \in C$ when $x \in C$, $\lambda > 0$.
- Axioms for ϱ we consider are:

  Monotonicity: $L_1 \leq L_2 \Rightarrow \varrho(L_1) \leq \varrho(L_2)$.

  Translation invariance: $\varrho(L + m) = \varrho(L) + m$ for all $m \in \mathbb{R}$.

  Subadditivity: $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$ for all $L_1, L_2 \in \mathcal{M}$.

  Positive homogeneity: $\varrho(\lambda L) = \lambda \varrho(L)$ for all $\lambda \geq 0$.

  Convexity: $\varrho(\gamma L_1 + (1-\gamma)L_2) \leq \gamma \varrho(L_1) + (1-\gamma)\varrho(L_2)$ for all $0 \leq \gamma \leq 1$, $L_1, L_2 \in \mathcal{M}$.

Definition 8.1 (Convex, coherent risk measures)

- A risk measure which satisfies monotonicity, translation invariance and convexity is called *convex*.
- A risk measure which satisfies monotonicity, translation invariance, subadditivity and positive homogeneity is called *coherent*.

A coherent risk measure is convex; the converse is not true (see below). On the other hand, for a positive-homogeneous risk measure, convexity and coherence are equivalent.

8.1.1 Risk measures and acceptance sets

Definition 8.2 (Acceptance set)

For a monotone and translation-invariant risk measure ϱ the *acceptance set of* ϱ is $A_\varrho = \{L \in \mathcal{M} : \varrho(L) \leq 0\}$ (so it contains the positions that are acceptable without any backing capital).

Proposition 8.3

Let ϱ be monotone and translation-invariant with associated A_ϱ. Then

1) $A_\varrho \neq \emptyset$ and A_ϱ satisfies

$$L \in A_\varrho \text{ and } \tilde{L} \leq L \Rightarrow \tilde{L} \in A_\varrho. \tag{25}$$

2) ϱ can be reconstructed from A_ϱ via

$$\varrho(L) = \inf\{m \in \mathbb{R} : L - m \in A_\varrho\}. \tag{26}$$

Proof. 1) is clear. For 2), note that $\inf\{m : L - m \in A_\varrho\} = \inf\{m : \varrho(L - m) \leq 0\} = \inf\{m : \varrho(L) - m \leq 0\}$ and this is equal to $\varrho(L)$. $\square$

Proposition 8.4

Suppose that A satisfies (25) and define

$$\varrho_A(L) = \inf\{m \in \mathbb{R} : L - m \in A\}. \tag{27}$$

Suppose $\varrho_A(L)$ is finite for all $L \in \mathcal{M}$. Then ϱ_A is monotone and translation-invariant on $\mathcal{M}$ and A_{ϱ_A} satisfies $A_{\varrho_A} \supseteq A$.

Proof. These properties of ϱ_A are easily checked. □

Example 8.5 (Value-at-risk)

For $\alpha \in (0,1)$, suppose we call $L \in \mathcal{M}$ *acceptable* if $\mathbb{P}(L > 0) \leq 1 - \alpha$. Then (27) is given by

$$\begin{aligned}\varrho_\alpha(L) &= \inf\{m \in \mathbb{R} : \mathbb{P}(L - m > 0) \leq 1 - \alpha\} \\ &= \inf\{m \in \mathbb{R} : \mathbb{P}(L \leq m) \geq \alpha\} = \mathrm{VaR}_\alpha(L).\end{aligned}$$

Proposition 8.6

1) Let ϱ be monotone and translation-invariant. Then

   1.1) ϱ is convex if and only if A_ϱ is convex.

   1.2) ϱ is coherent if and only if A_ϱ is a convex cone.

2) More generally, consider a set of acceptable positions A and the associated risk measure ϱ_A (whose acceptance set may be larger than A). If A is convex, so is ϱ_A; if A is a convex cone, then ϱ_A is coherent.

Example 8.7 (Risk measures based on loss functions)

Consider a strictly increasing and convex *loss function* $\ell : \mathbb{R} \to \mathbb{R}$ and some $c \in \mathbb{R}$. Assume that $\mathbb{E}(\ell(L))$ is finite for all $L \in \mathcal{M}$. Define an acceptance set by

$$A = \{L \in \mathcal{M} : \mathbb{E}\big(\ell(L)\big) \leq \ell(c)\},$$

and the associated risk measure by

$$\varrho_A = \inf\{m \in \mathbb{R} : \mathbb{E}\big(\ell(L-m)\big) \leq \ell(c)\}.$$

- ϱ_A is translation invariant and monotone by Proposition 8.4 since A satisfies (25).
- ϱ_A is convex by Proposition 8.6; to see this consider acceptable positions L_1 and L_2 and observe that the convexity of ℓ implies

$$\begin{aligned}\mathbb{E}(\ell(\gamma L_1 + (1-\gamma)L_2)) &\leq \mathbb{E}(\gamma\ell(L_1) + (1-\gamma)\ell(L_2)) \\ &\leq \gamma\ell(c) + (1-\gamma)\ell(c) = \ell(c),\end{aligned}$$

  where we have used that $\mathbb{E}(\ell(L_i)) \leq \ell(c)$ for acceptable positions. Hence $\gamma L_1 + (1-\gamma)L_2 \in A$, so A is convex.
- Example: $\ell(x) = \exp(\alpha x)$ for some $\alpha > 0$. Then

$$\begin{aligned}\varrho_{\alpha,c}(L) &:= \inf\{m : \mathbb{E}(e^{\alpha(L-m)}) \leq e^{\alpha c}\} = \inf\{m : \mathbb{E}(e^{\alpha L}) \leq e^{\alpha c + \alpha m}\} \\ &= \frac{1}{\alpha}\log(\mathbb{E}(e^{\alpha L})) - c.\end{aligned}$$

  Note that $\varrho_{\alpha,c}(0) = -c$, so $\varrho_{\alpha,c}$ cannot be coherent. For $c = 0$ and

$\lambda > 1$, the *entropic risk measure* $\varrho_{\alpha,0}$ satisfies

$$\varrho_{\alpha,0}(\lambda L) = \frac{1}{\alpha} \ln\{\mathbb{E}(e^{\alpha\lambda L})\} \geq \frac{1}{\alpha} \ln\{\mathbb{E}(e^{\alpha L})^{\lambda}\} = \lambda \varrho_{\alpha,0}(L),$$

where the inequality is strict if L is non-degenerate. This shows that $\varrho_{\alpha,0}$ is convex but not coherent. If L are insurance claims, $\varrho_{\alpha,0}$ is known as *exponential premium principle*.

Example 8.8 (Stress test or worst case risk measure)

Given *stress scenarios* $S \subseteq \Omega$, a *stress test risk measure* can be defined by

$$\varrho(L) = \sup\{L(\omega) : \omega \in S\},$$

that is, the worst loss on S. The associated acceptance set is

$$A_\varrho = \{L : L(\omega) \leq 0 \text{ for all } \omega \in S\}.$$

The choice of S is often guided by the underlying probability measure $\mathbb{P}$.

Example 8.9 (Generalized scenario risk measures)
Consider a set $\mathcal{Q}$ of probability measures on $(\Omega, \mathcal{F})$ and a *penalty function* $\gamma : \mathcal{Q} \to \mathbb{R}$ such that $\inf\{\gamma(\mathbb{Q}) : \mathbb{Q} \in \mathcal{Q}\} > -\infty$. Suppose $\sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}|L| < \infty$ for all $L \in \mathcal{M}$. The *generalized scenario risk measures* ϱ is defined by

$$\varrho(L) = \sup\{\mathbb{E}_{\mathbb{Q}}(L) - \gamma(\mathbb{Q}) : \mathbb{Q} \in \mathcal{Q}\}. \tag{28}$$

The corresponding acceptance set is given by

$$A_\varrho = \{L \in \mathcal{M} \colon \sup\{\mathbb{E}_{\mathbb{Q}}(L) - \gamma(\mathbb{Q}) : \mathbb{Q} \in \mathcal{Q}\} \leq 0\}.$$

- A_ϱ is convex, and thus so is ϱ.
- Every convex risk measure can be represented as (28); see Theorem 8.10.
- If $\gamma(\cdot) \equiv 0$ on $\mathcal{Q}$, ϱ is positive homogeneous and therefore coherent.
- The stress test risk measure of Example 8.8 is a special case of (28) in which $\gamma \equiv 0$ and $\mathcal{Q}$ is the set of all Dirac measures $\delta_\omega(\cdot),\ \omega \in S$, that is, $\delta_\omega(B) = I_B(\omega)$ for arbitrary measurable sets $B \subseteq \Omega$.

8.1.2 Dual representation of convex measures of risk

Theorem 8.10 (Dual representation for risk measures)
Suppose $|\Omega| = n < \infty$. Let $\mathcal{F} = \mathcal{P}(\Omega)$ (power set) and $\mathcal{M} := \{L : \Omega \to \mathbb{R}\}$. Then:

1) Every convex risk measure ϱ on $\mathcal{M}$ can be written in the form

$$\varrho(L) = \max\{\mathbb{E}_{\mathbb{Q}}(L) - \alpha_{\min}(\mathbb{Q}) : \mathbb{Q} \in \mathcal{S}^1(\Omega, \mathcal{F})\}, \qquad (29)$$

where $\mathcal{S}^1(\Omega, \mathcal{F})$ denotes the set of all probability measures on Ω, and where the penalty function $\alpha_{\min}$ is given by $\alpha_{\min}(\mathbb{Q}) = \sup\{\mathbb{E}_{\mathbb{Q}}(L) : L \in A_\varrho\}$.

2) If ϱ is coherent, it has the representation

$$\varrho(L) = \max\{\mathbb{E}_{\mathbb{Q}}(L) \colon \mathbb{Q} \in \mathcal{Q}\}$$

for some set $\mathcal{Q} = \mathcal{Q}(\varrho) \subseteq \mathcal{S}^1(\Omega, \mathcal{F})$.

One can show that $\alpha_{\min}(\mathbb{Q}) = \sup_{L \in \mathcal{M}}\{\mathbb{E}_{\mathbb{Q}}(L) - \varrho(L)\}$.

8.1.3 Examples of dual representations

Proposition 8.11 (ES formulas)
For $\alpha \in (0,1)$,

1) $\mathrm{ES}_\alpha(L) = \dfrac{\mathbb{E}\big((L - F_L^{\leftarrow}(\alpha))_+\big)}{1-\alpha} + F_L^{\leftarrow}(\alpha)$;

2) $\mathrm{ES}_\alpha(L) = \dfrac{\mathbb{E}(L I_{\{L > F_L^{\leftarrow}(\alpha)\}}) + F_L^{\leftarrow}(\alpha)\big(1 - \alpha - \bar{F}_L(F_L^{\leftarrow}(\alpha))\big)}{1-\alpha}$.

Corollary 8.12 (ES formulas under continuous F_L)
Let F_L be continuous at $F_L^{\leftarrow}(\alpha)$. Then

1) $\mathrm{ES}_\alpha(L) = \dfrac{\mathbb{E}(L I_{\{L > F_L^{\leftarrow}(\alpha)\}})}{1-\alpha}$

2) $\mathrm{ES}_\alpha(L) = \mathbb{E}(L \,|\, L > F_L^{\leftarrow}(\alpha))$ (i.e. *conditional VaR (CVaR)*)

With dual representations one can give a proof for ES_α being subadditive; see the following result.

Theorem 8.13

For $\alpha \in [0,1)$, ES_α is coherent on $\mathcal{M} = \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. The dual representation is given by

$$\mathrm{ES}_\alpha(L) = \max\{\mathbb{E}^{\mathbb{Q}}(L) : \mathbb{Q} \in \mathcal{Q}_\alpha\}, \tag{30}$$

where $\mathcal{Q}_\alpha$ is the set of all probability measures on $(\Omega, \mathcal{F})$ that are absolutely continuous with respect to $\mathbb{P}$ and for which the measure-theoretic density $d\mathbb{Q}/d\mathbb{P}$ is bounded by $1/(1-\alpha)$.

8.2 Law-invariant coherent risk measures

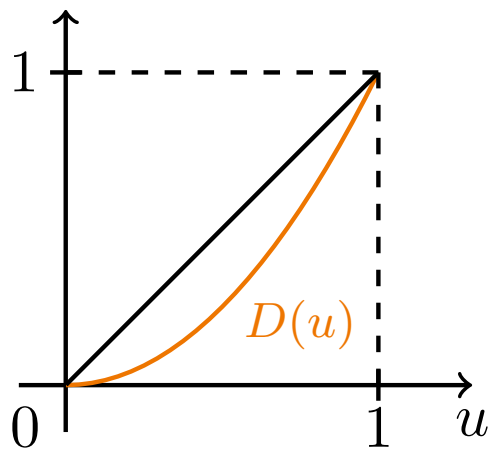
8.2.1 Distortion risk measures

Distortion risk measures are important coherent risk measures. We summarize important representations and investigate their properties.

Representations of distortion risk measures

Definition 8.14 (Distortion risk measure)

A *convex distortion function* D is a convex, increasing and absolutely continuous function on $[0,1]$ satisfying $D(0) = 0$ and $D(1) = 1$.



The *distortion risk measure* associated with D is defined by

$$\varrho(L) = \int_0^1 F_L^{\leftarrow}(u)\,\mathrm{d}D(u). \tag{31}$$

Note:

- A distortion risk measure is law-invariant (average of the L-quantiles).
- $D(u) = \int_0^u \phi(s)\,\mathrm{d}s$ for an increasing, positive function ϕ (the right-sided derivative of D), hence

$$\varrho(L) = \int_0^1 F_L^{\leftarrow}(u)\phi(u)\,\mathrm{d}u. \tag{32}$$

  A risk measure of this form is known as *spectral risk measure* and ϕ as *spectrum*.
- For $D_\alpha(u) = (1-\alpha)^{-1}(u-\alpha)^+$ one obtains expected shortfall. The spectrum is $\phi(u) = (1-\alpha)^{-1} I_{\{u \geq \alpha\}}$ (equal weight is placed on all quantiles beyond the α-quantile).

Lemma 8.15

The distortion risk measure ϱ associated with a convex distortion function D can be written in the form

$$\varrho(L) = \int_{\mathbb{R}} x \, \mathrm{d}D \circ F_L(x), \tag{33}$$

where $D \circ F_L(x) = D\big(F_L(x)\big)$.

Proof. $G(x) = D \circ F_L(x)$ has quantile function $G^{\leftarrow} = F_L^{\leftarrow} \circ D^{\leftarrow}$. Thus (33) can be written as

$$\int_{\mathbb{R}} x \, \mathrm{d}G(x) \underset{u=G(x)}{=} \int_0^1 G^{\leftarrow}(u) \, \mathrm{d}u = \int_0^1 F_L^{\leftarrow} \circ D^{\leftarrow}(u) \, \mathrm{d}u = \mathbb{E}(F_L^{\leftarrow} \circ D^{\leftarrow}(U)),$$

where $U \sim \mathrm{U}(0,1)$. Now introduce $V = D^{\leftarrow}(U) \sim D$ and note that

$$\int_{\mathbb{R}} x \, \mathrm{d}D \circ F_L(x) = \mathbb{E}(F_L^{\leftarrow}(V)) = \int_0^1 F_L^{\leftarrow}(v) \, \mathrm{d}D(v). \qquad \square$$

D distorts F_L. Since D is convex, $D(u) \leq u$, so $G = D \circ F_L$ puts more mass on high values of L than F_L.

Distortion risk measure can be represented as a weighted average of expected shortfall; see the appendix for a proof.

Proposition 8.16 (Distortion risk measures as weighted ES)
Let ϱ be a distortion risk measure associated with the convex distortion function D. Then, for a probability measure μ,

$$\varrho(L) = \int_0^1 \mathrm{ES}_\alpha(L)\,\mathrm{d}\mu(\alpha).$$

Properties of distortion risk measures

Definition 8.17 (Comonotone additivity)
A risk measure ϱ on a space of random variables $\mathcal{M}$ is said to be *comonotone additive* if $\varrho(L_1 + \cdots + L_d) = \varrho(L_1) + \cdots + \varrho(L_d)$ for comonotone $L_1, \ldots, L_d$.

- Quantile functions (so value-at-risk) are comonotone additive. Comonotone additivity of distortion risk measures then follows from (31).

- Distortion risk measures are coherent. Monotonicity, translation invariance and positive homogeneity are obvious. Subadditivity follows from Proposition 8.16 and subadditivity of ES_α (e.g., Theorem 8.13) by observing that

$$\begin{aligned}\varrho(L_1 + L_2) &= \int_0^1 \mathrm{ES}_\alpha(L_1 + L_2)\,\mathrm{d}\mu(\alpha) \\ &\leq \int_0^1 \mathrm{ES}_\alpha(L_1)\,\mathrm{d}\mu(\alpha) + \int_0^1 \mathrm{ES}_\alpha(L_2)\,\mathrm{d}\mu(\alpha) \\ &= \varrho(L_1) + \varrho(L_2).\end{aligned}$$

- In summary, we have verified that distortion risk measures are law invariant, coherent and comonotone additive.
- It may also be shown that, on an atomless probability space (where there exists a continuous random variable), a law-invariant, coherent, comonotone-additive risk measure must be of the form (31) for some convex distortion function D.

- Parametric families of distortion risk measures can be based on convex distortion functions of the form

$$D_\alpha(u) = \Psi(\Psi^{-1}(u) + \ln(1-\alpha)), \quad 0 \le \alpha < 1,$$

where Ψ is a continuous df on $\mathbb{R}$; for $\Psi(u) = 1 - \exp(-u)$, $u \ge 0$, one obtains the distortion function for ES.

  - Such a family of convex distortion functions is strictly decreasing in α for fixed u.
  - $D_0(u) = u$ (corresponding to the risk measure $\varrho(L) = \mathbb{E}(L)$) and $\lim_{\alpha \to 1} D(u) = 1_{\{u=1\}}$.
  - For $\alpha_1 < \alpha_2$ and $0 < u < 1$ we have $D_{\alpha_1}(u) > D_{\alpha_2}(u)$, so that D_{α_2} distorts the original probability measure more than D_{α_1} and places more weight on outcomes in the tail.

8.2.2 The expectile risk measure

Definition 8.18 (Expectiles)
Let $L \in \mathcal{M} := L^1(\Omega, \mathcal{F}, \mathbb{P})$, so $\mathbb{E}|L| < \infty$. Then, for $\alpha \in (0,1)$, the *α-expectile* $e_\alpha(L)$ is given by the unique solution y of

$$\alpha \mathbb{E}((L-y)^+) = (1-\alpha)\mathbb{E}((L-y)^-) \tag{34}$$

where $x^+ = \max\{x, 0\}$ and $x^- = \max\{-x, 0\}$.

- Since $x^+ - x^- = x$, $e_{0.5}(L) = \mathbb{E}(L)$ as $\mathbb{E}(L-y)^- = \mathbb{E}(L-y)^+$ iff $\mathbb{E}((L-y)^+ - (L-y)^-) = 0$ iff $\mathbb{E}(L-y) = 0$.
- $\mathbb{E}(L^2) < \infty$, $e_\alpha(L)$ is the minimizer of

$$\min_{y \in \mathbb{R}} \mathbb{E}\left(S(y, L)\right) \tag{35}$$

for *scoring function* $S(y, L)$. This could be relevant for the out-of-sample testing of expectile-estimates (so-called *backtesting*). The scoring func-

tion that yields the expectile is

$$S^e_\alpha(y, L) = |1_{\{L \le y\}} - \alpha|(L - y)^2. \tag{36}$$

In fact we can compute that $\frac{d}{dy}\mathbb{E}\left(S^e_\alpha(y, L)\right)$ equals

$$\begin{aligned}
&\frac{d}{dy}\int_{-\infty}^{\infty} |1_{\{y \ge x\}} - \alpha|(y - x)^2 \,\mathrm{d}F_L(x) \\
&= \frac{d}{dy}\int_{-\infty}^{y} (1 - \alpha)(y - x)^2 \,\mathrm{d}F_L(x) + \frac{d}{dy}\int_{y}^{\infty} \alpha(y - x)^2 \,\mathrm{d}F_L(x) \\
&= 2(1 - \alpha)\int_{-\infty}^{y} (y - x) \,\mathrm{d}F_L(x) + 2\alpha\int_{y}^{\infty} (y - x) \,\mathrm{d}F_L(x) \\
&= 2(1 - \alpha)\mathbb{E}((L - y)^-) - 2\alpha\mathbb{E}((L - y)^+)
\end{aligned}$$

and setting this equal to zero yields the definition of an expectile.

- One can show that the α-quantile $F^{\leftarrow}_L(\alpha)$ is also a minimizer of the form (35); consider the scoring function $S^q_\alpha(y, L) = |1_{\{L \le y\}} - \alpha||L - y|$.

The following result shows uniqueness of the α-expectile and provides a helpful formula for computing expectiles of certain distributions; see the appendix for a proof.

Proposition 8.19

Let $\alpha \in (0,1)$ and L a rv such that $\mu := \mathbb{E}(L) < \infty$. Then $e_\alpha(L)$ may be written as $e_\alpha(L) = \tilde{F}_L^{-1}(\alpha)$ where

$$\tilde{F}_L(y) = \frac{yF_L(y) - \mu(y)}{2(yF_L(y) - \mu(y)) + \mu - y} \tag{37}$$

is a continuous df that is strictly increasing on its support and $\mu(y) := \int_{-\infty}^{y} x \, \mathrm{d}F_L(x)$.

Example 8.20 (Bernoulli)

Let $L \sim \text{Be}(p)$ be a Bernoulli-distributed loss. Then

$$F_L(y) = \begin{cases} 0, & y < 0 \\ 1-p, & 0 \le y < 1, \\ 1, & y \ge 1 \end{cases} \quad \mu(y) = \begin{cases} 0, & y < 1 \\ p, & y \ge 1 \end{cases}$$

from which it follows that $\tilde{F}_L(y) = \frac{y(1-p)}{y(1-2p)+p}$, $0 \le y \le 1$ and

$$e_\alpha(L) = \frac{\alpha p}{(1-\alpha) + p(2\alpha - 1)}.$$

Note that this can take any value in zero and one, whereas $\text{VaR}_\alpha(L) \in \{0, 1\}$, $\alpha \in (0, 1]$.

Properties of expectiles

Proposition 8.21 (Coherence of expectile risk measures)
$\varrho = e_\alpha$ is a coherent risk measure on $\mathcal{M} = L^1(\Omega, \mathcal{F}, \mathbb{P})$ for $\alpha \geq 0.5$.

- See the appendix for a proof.
- Expectiles are not comonotone additive and thus are not distortion risk measures.
- If L_1 and L_2 are comonotonic and of the same type (so that $L_2 = kL_1 + m$ for some $m \in \mathbb{R}$ and $k > 0$) then we do have comonotone additivity (by translation invariance and positive homogeneity), but for comonotonic variables that are not of the same type one can find examples where $e_\alpha(L_1 + L_2) < e_\alpha(L_1) + e_\alpha(L_2)$ for $\alpha > 0.5$.

Elicitability explained in words

- Computing a (one-period ahead) risk measure $\varrho(L) =: \varrho(F_L)$ is a point forecasting problem because F_L is unknown and one has to find an estimate $\hat{F}_L$ of it and forecast the unknown true $\varrho(F_L)$ via the point forecast $\varrho(\hat{F}_L)$.
- As different $\hat{F}_L$ can be used to forecast the risk measure, it is desirable to be able to evaluate which of them gives a better point forecast.
- Suppose we want to forecast L (or F_L) by a point y. The *forecasting error* is

$$\mathbb{E}(S(y, L)) = \int_{\mathbb{R}} S(y, l) \, \mathrm{d}F_L(l),$$

  where $S(y, l)$ is a *scoring* (i.e. forecasting objective) *function*.
- Two point forecasting methods can be compared via their forecasting

errors. For a given S, the optimal point forecast is

$$\varrho^*(F_L) = \operatorname*{arginf}_y \mathbb{E}(S(y, L)) \quad \text{(minimizing the forecast error)}.$$

For example, for $S(y, l) = (y - l)^2$ and $S(y, l) = |y - l|$, the optimal point forecasts are the mean and median of F_L, respectively.

- *Elicitable risk measures* (or: statistical functionals) are risk measures ϱ which minimize $\mathbb{E}(S(y, L))$ of some scoring function S; hence that S can be used to compare different point forecasting procedures for ϱ ("the smaller the forecasting error, the better" makes sense).
- If ϱ is not elicitable, one cannot find such an S and thus the minimization of the forecasting error does not yield the true value $\varrho(F_L)$ for any S. Hence, for two competing point forecasts of $\varrho(F_L)$, one cannot tell which performs the best by comparing their forecasting error, no matter what S is used.

8.3 Risk measures for linear portfolios

We now consider linear portfolios in

$$\mathcal{M} = \{L : L = m + \boldsymbol{\lambda}' \boldsymbol{X},\ m \in \mathbb{R}, \boldsymbol{\lambda} \in \mathbb{R}^d\}, \tag{38}$$

for a fixed d-dimensional random vector $\boldsymbol{X}$.

- Many standard approaches to risk aggregation and capital allocation are based on the assumption that losses have a linear relationship to underlying risk factor changes.
- It is common to use linear approximations for losses due to market risks over short time horizons.

8.3.1 Elliptically distributed risk factors

Theorem 8.22 (Risk measurement for elliptical risk factors)
Let $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ and ϱ be any positive-homogeneous, translation-invariant and law-invariant risk measure on $\mathcal{M}$. Then:

1) For any $L = m + \boldsymbol{\lambda}'\boldsymbol{X} \in \mathcal{M}$, $\varrho(L) = m + \boldsymbol{\lambda}'\boldsymbol{\mu} + \sqrt{\boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda}}\varrho(Y_1)$ for $Y_1 \sim S_1(\psi)$.

2) If $\varrho(Y_1) \geq 0$, then ϱ is subadditive on $\mathcal{M}$ (e.g., VaR_α for $\alpha \geq 0.5$).

3) If $\mathbb{E}\boldsymbol{X}$ exists then, $\forall\, L = m + \boldsymbol{\lambda}'\boldsymbol{X} \in \mathcal{M}$ and $\rho_{ij} = \wp(\Sigma)_{ij} = P_{ij}$,

$$\varrho(L - \mathbb{E}L) = \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d} \rho_{ij}\lambda_i\lambda_j\varrho(X_i - \mathbb{E}X_i)\varrho(X_j - \mathbb{E}X_j)}.$$

4) If $\mathrm{cov}(\boldsymbol{X})$ exists and $\varrho(Y_1) > 0$ then, for every $L \in \mathcal{M}$, $\varrho(L) = \mathbb{E}(L) + k_\varrho\sqrt{\mathrm{var}(L)}$ for some $k_\varrho > 0$ depending on ϱ.

Proof.

1) Let $\boldsymbol{Y} \sim S_k(\psi)$, $AA' = \Sigma$. $L = m + \boldsymbol{\lambda}'\boldsymbol{X} \overset{\mathrm{d}}{=} m + \boldsymbol{\lambda}'\boldsymbol{\mu} + \boldsymbol{\lambda}' A \boldsymbol{Y}$. By Theorem 6.11 3), $L \overset{\mathrm{d}}{=} m + \boldsymbol{\lambda}'\boldsymbol{\mu} + \|A'\boldsymbol{\lambda}\| Y_1$. Thus $\varrho(L) = m + \boldsymbol{\lambda}'\boldsymbol{\mu} + \|A'\boldsymbol{\lambda}\|\varrho(Y_1) = m + \boldsymbol{\lambda}'\boldsymbol{\mu} + \sqrt{\boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda}}\varrho(Y_1)$.
2) Set $L_1 = m_1 + \boldsymbol{\lambda}_1'\boldsymbol{X}$ and $L_2 = m_2 + \boldsymbol{\lambda}_2'\boldsymbol{X}$. Subadditivity follows from 1) and $\|A'(\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2)\| \leq \|A'\boldsymbol{\lambda}_1\| + \|A'\boldsymbol{\lambda}_2\|$ and $\varrho(Y_1) \geq 0$.
3) $\varrho(L - \mathbb{E}L) = \varrho(L) - \mathbb{E}(L) = \varrho(L) - (m + \boldsymbol{\lambda}'\boldsymbol{\mu}) = \sqrt{\boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda}}\varrho(Y_1)$, so

$$\varrho(L - \mathbb{E}L) = \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d} \rho_{ij}\lambda_i\lambda_j\sigma_i\sigma_j}\varrho(Y_1),$$

where $\sigma_j = \sqrt{\Sigma_{jj}}$ for $j \in \{1, \ldots, d\}$. For $\boldsymbol{\lambda} = \boldsymbol{e}_j$, $\varrho(X_j - \mathbb{E}X_j) = \varrho(\boldsymbol{e}_j'\boldsymbol{X} - \mathbb{E}(\boldsymbol{e}_j'\boldsymbol{X})) = \sigma_j\varrho(Y_1)$, from which the result follows.
4) $\mathrm{cov}(\boldsymbol{X}) = c\Sigma$ for some $c > 0$. Since $\mathrm{var}(L) = \mathrm{var}(\boldsymbol{\lambda}'\boldsymbol{X}) = \boldsymbol{\lambda}' c\Sigma\boldsymbol{\lambda}$, 3) implies $\varrho(L) = \mathbb{E}(L) + \sqrt{\boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda}}\varrho(Y_1) = \mathbb{E}(L) + \sqrt{\mathrm{var}(L)}\varrho(Y_1)/\sqrt{c}$. □

- 2) gives a special case where VaR is subadditive and thus coherent. In particular, if $(L_1, \ldots, L_d)$ is jointly elliptical, VaR_α is subadditive for $\alpha \geq 0.5$.
- 3) provides a useful interpretation of risk measures on $\mathcal{M}$ in terms of the aggregation of stress tests.
- 4) is relevant to portfolio optimization. If we consider losses $L \in \mathcal{M}$ for which $\mathbb{E}(L)$ is fixed, the weights that minimize ϱ also minimize the variance. The portfolio minimizing ϱ is thus the same as the Markowitz variance-minimizing portfolio.

8.4 Risk aggregation

- A *risk aggregation rule* is a mapping

$$f(\mathrm{EC}_1, \ldots, \mathrm{EC}_d) = \mathrm{EC}$$

  which maps the individual capital amounts $\mathrm{EC}_1, \ldots, \mathrm{EC}_d$ to the aggregate capital EC (economic capital). Examples are:

  - *Simple summation* $\mathrm{EC} = \mathrm{EC}_1 + \cdots + \mathrm{EC}_d$ (a special case of and upper bound for correlation adjusted summation)
  - *Correlation adjusted summation*

$$\mathrm{EC} = \sqrt{\sum_{i=1}^{d} \sum_{j=1}^{d} \rho_{ij} \, \mathrm{EC}_i \, \mathrm{EC}_j}, \tag{39}$$

    where $\rho_{ij} \in [0, 1]$ are parameters (referred to as *correlations*).

- In what follows we show that correlation adjusted summation is justified as a risk aggregation rule under various setups.

8.4.1 Aggregation based on loss distributions

- Suppose that the overall loss is $L = L_1 + \cdots + L_d$ where $L_1, \ldots, L_d$ are the losses arising from sub-units (e.g., business units, asset classes). Consider a translation-invariant ϱ and define

$$\varrho^{\text{mean}}(\cdot) = \varrho(\cdot - \mathbb{E}(\cdot)) = \varrho(\cdot) - \mathbb{E}(\cdot),$$

  that is, the capital required to cover unexpected losses.

- The capital requirements for the sub-units are

$$\text{EC}_j = \varrho^{\text{mean}}(L_j), \quad j \in \{1, \ldots, d\},$$

  and the aggregate capital should be

$$\text{EC} = \varrho^{\text{mean}}(L).$$

- We require an aggregation rule f such that $f(\text{EC}_1, \ldots, \text{EC}_d) = \text{EC}$.

- If $\varrho(L) = \mathbb{E}(L) + k\,\mathrm{sd}(L)$, $k > 0$, and $\mathbb{E}(L^2) < \infty$ then

$$\mathrm{sd}(L) = \sqrt{\mathrm{var}(\mathbf{1}'\boldsymbol{L})} = \sqrt{\mathbf{1}'\,\mathrm{cov}(\boldsymbol{L})\mathbf{1}} = \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d}\rho_{ij}\,\mathrm{sd}(L_i)\,\mathrm{sd}(L_j)},$$

  where $(\rho_{ij})_{i,j} = \mathrm{corr}(\boldsymbol{L})$, so correlation adjusted summation follows by noting that $\mathrm{sd}(L) = \varrho^{\mathrm{mean}}(L)/k = \mathrm{EC}\,/k$ (and $\mathrm{sd}(L_j) = \mathrm{EC}_j\,/k$).

- If $L_j = m_j + \boldsymbol{\lambda}_j'\boldsymbol{X}$ for $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ with existing $\mathrm{cov}(\boldsymbol{X})$, then this formula and Theorem 8.22 4) imply that correlation adjusted summation is justified for any positive-homogeneous, translation-invariant and law-invariant risk measure ϱ.

- The assumption on $\mathrm{cov}(\boldsymbol{X})$ can be dropped.

8.4.2 Aggregation based on stressing risk factors

- Correlation adjusted summation is used in the aggregation of capital contributions $\mathrm{EC}_1, \ldots, \mathrm{EC}_d$ computed by stressing individual risk factors (example: Standard formula approach to Solvency II).
- Let $\boldsymbol{x} = \boldsymbol{X}(\omega)$ be a scenario defined in terms of changes in risk factors and $L(\boldsymbol{x})$ the corresponding loss. Assume $L(\boldsymbol{x})$ is known and componentwise increasing.
- The d risk factors are stressed individually by amounts $k_1, \ldots, k_d$. Capital contributions for each risk factor are computed by

$$\mathrm{EC}_j = L(k_j \boldsymbol{e}_j) - L(\mathbb{E}(X_j)\boldsymbol{e}_j)$$

  where $k_j > \mathbb{E}(X_j)$ so that $\mathrm{EC}_j > 0$ (interpreted as the loss incurred by stressing risk factor j by k_j relative to the impact of stressing it by its expected change); an example is $k_j = q_\alpha(X_j)$ for large α.
- The following justifies correlation adjusted summation as a risk aggregation rule if $k_j = \varrho(X_j)$ for elliptical $\boldsymbol{X}$ and $L(\boldsymbol{X}) = m + \boldsymbol{\lambda}'\boldsymbol{X}$.

Proposition 8.23 (Justification for correlation adjusted summation) Let $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ with $\mathbb{E}(\boldsymbol{X}) = \boldsymbol{\mu}$. Let $\mathcal{M}$ be the space of linear portfolios (38) and ϱ be a pos. hom., translation- and law-invariant risk measure on $\mathcal{M}$. Then, for any $L = L(\boldsymbol{X}) = m + \boldsymbol{\lambda}'\boldsymbol{X} \in \mathcal{M}$,

$$\mathrm{EC} = \varrho(L - \mathbb{E}(L)) = \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d} \rho_{ij}\,\mathrm{EC}_i\,\mathrm{EC}_j},$$

where $\mathrm{EC}_j = L(\varrho(X_j)\boldsymbol{e}_j) - L(\mathbb{E}(X_j)\boldsymbol{e}_j)$ and $\rho_{ij} = \wp(\Sigma)_{i,j}$.

Proof. Note that $\mathrm{EC}_j = m + \lambda_j \varrho(X_j) - (m + \lambda_j \mathbb{E}X_j) = \lambda_j \varrho(X_j - \mathbb{E}X_j)$ and plug this into Theorem 8.22 3) to see that the claim holds. □

- Thus under linearity of the losses in jointly elliptical risk-factor changes, we can aggregate the effects of single-risk-factor stresses to an aggregate capital; this applies to VaR, ES. This idea underscores correlation adjusted summation in Solvency II.
- The ρ_{ij}s are either estimated (if possible) or set by expert judgement.

8.4.3 Risk aggregation and Fréchet problems

- Consider the *margins-plus-copula approach* where $L_j \sim F_j$, $j \in \{1, \dots, d\}$, are treated as known (estimated or postulated) and C is unknown.
- Consider $L = L_1 + \cdots + L_d$. Due to the unknown C (*dependence uncertainty*), risk measures can no longer be computed explicitly.
- Our goal is to find bounds on VaR_α and ES_α under all possible C. Let

$$\mathcal{S}_d := \mathcal{S}_d(F_1, \dots, F_d) := \left\{ L = \sum_{j=1}^{d} L_j : L_j \sim F_j,\ j = 1, \dots, d \right\}$$

and consider

$$\overline{\varrho}(L) := \overline{\varrho}(\mathcal{S}_d) := \sup\{\varrho(L) : L \in \mathcal{S}_d(F_1, \dots, F_d)\} \quad \text{(worst } \varrho\text{)}$$
$$\underline{\varrho}(L) := \underline{\varrho}(\mathcal{S}_d) := \inf\{\varrho(L) : L \in \mathcal{S}_d(F_1, \dots, F_d)\} \quad \text{(best } \varrho\text{)}$$

- If $\varrho = \mathrm{ES}_\alpha$, $\overline{\mathrm{ES}}_\alpha(L) = \sum_{j=1}^{d} \mathrm{ES}_\alpha(L_j)$ (subadditivity, com. additivity). $\underline{\mathrm{ES}}_\alpha$, $\underline{\mathrm{VaR}}_\alpha$, $\overline{\mathrm{VaR}}_\alpha$ depend on whether the portfolio is *homogeneous* (that is, $F_1 = \cdots = F_d$); we focus on $\overline{\mathrm{VaR}}_\alpha$.

Summary of existing results

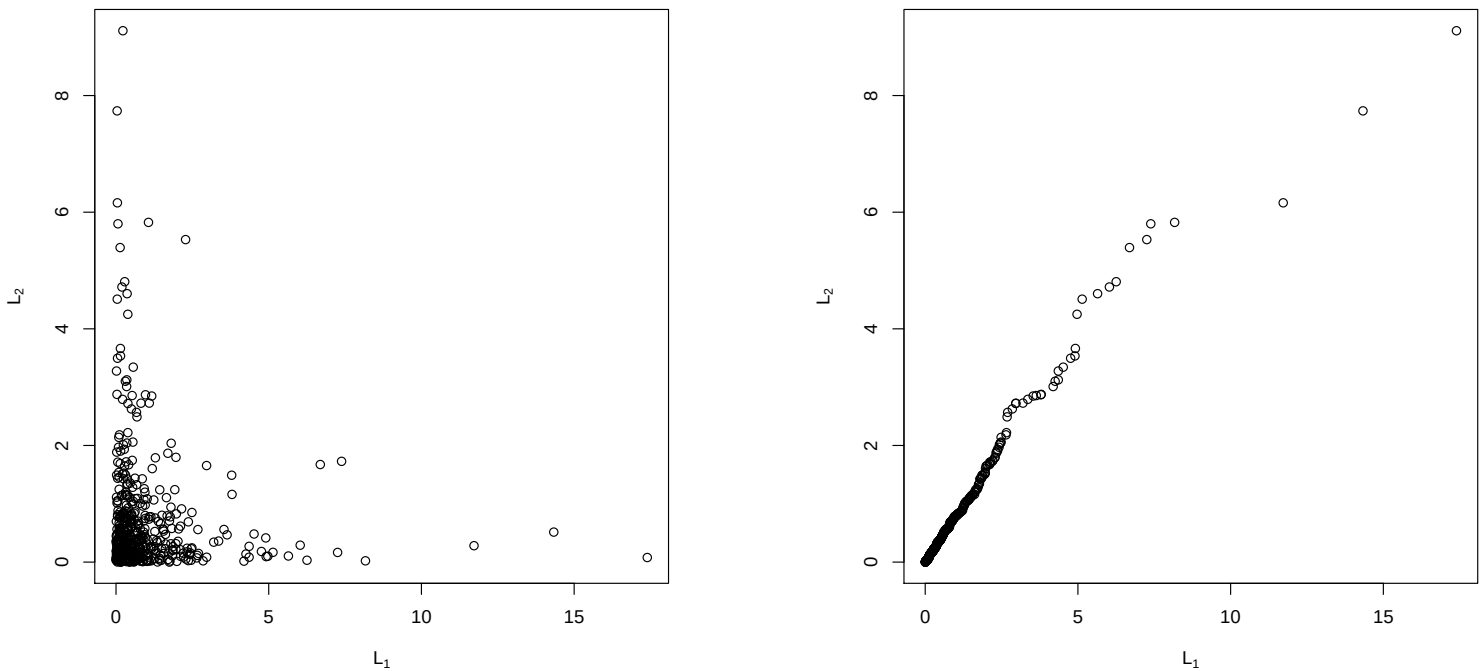
$d = 2$: Fully solved analytically

$d \geq 3$: Here we distinguish:

- Homogeneous case ($F_1 = \cdots = F_d$):
  - $\underline{\mathrm{ES}}_\alpha(L)$ solved analytically for decreasing densities (e.g. Pareto, Exponential)
  - $\underline{\mathrm{VaR}}_\alpha(L)$, $\overline{\mathrm{VaR}}_\alpha(L)$ solved analytically for tail-decreasing densities (e.g. Pareto, Log-normal, Gamma)
- Inhomogeneous case:
  - Some analytical results available
  - Numerical methods: (Adaptive/Block) Rearrangement Algorithm

The general problem

- We have one-period risks $L_1 \sim F_1, \ldots, L_d \sim F_d$ with given (estimated or postulated) $F_1, \ldots, F_d$ and unknown copula C and want to compute $\overline{\mathrm{VaR}}_\alpha(L)$ for $L = L_1 + \cdots + L_d$.
- Iman and Conover (1982) idea for $\mathrm{Par}(2)$, $\mathrm{Par}(2.5)$ sample of size 500:



$\Rightarrow$ Reordering columns changes the dependence of (L_1, L_2) and F_L.

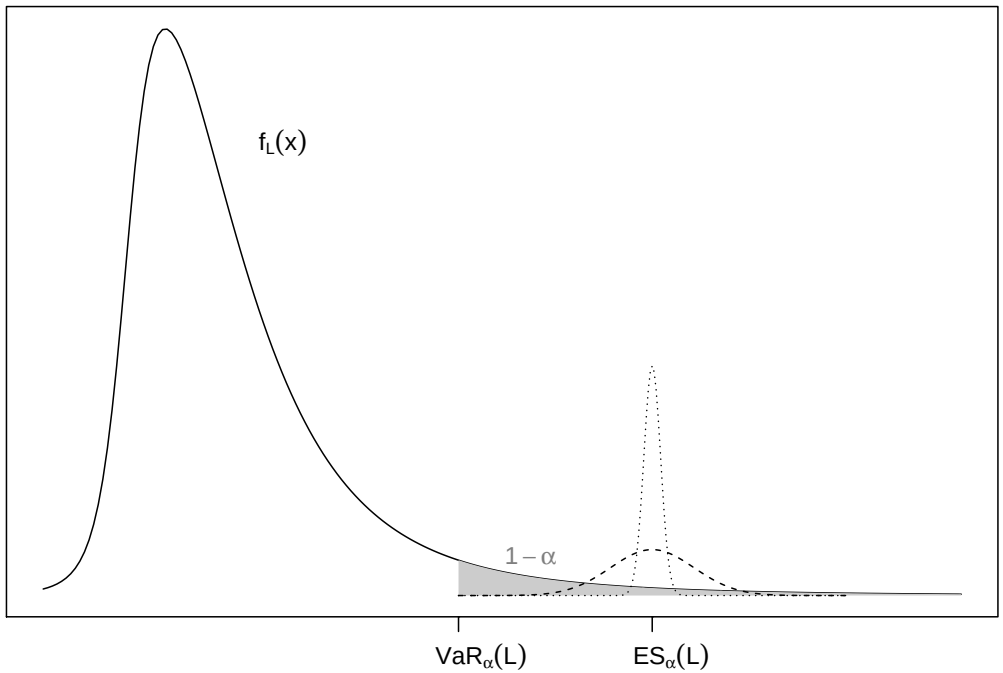
The Rearrangement Algorithm (RA)

- Two columns $\boldsymbol{a}, \boldsymbol{b}$ are *oppositely ordered* if $(a_i - a_j)(b_i - b_j) \leq 0 \; \forall \, i, j$.
- *Minimum row-sum operator* $s(X) = \min_{1 \leq i \leq N} \sum_{1 \leq j \leq d} x_{ij}$

Algorithm 8.24 (RA for computing $\overline{\mathrm{VaR}}_\alpha(L)$)

1) Fix $\alpha \in (0,1)$, $F_1^{\leftarrow}, \dots, F_d^{\leftarrow}$, $N \in \mathbb{N}$ (# of discr. points), $\varepsilon \geq 0$ (tol.)
2) Compute the lower bound $\underline{s}_N$:
   - 2.1) Define the (N,d)-matrix $\underline{X}^\alpha = \left(F_j^{\leftarrow}(\alpha + \frac{(1-\alpha)(i-1)}{N}) \right)_{i,j}$.
   - 2.2) Randomly permute each column of $\underline{X}^\alpha$ (to avoid $\overline{s}_N - \underline{s}_N \nrightarrow 0$)
   - 2.3) Iterate over all columns of $\underline{X}^\alpha$ and oppositely order each to the sum of all others $\Rightarrow$ Matrix $\underline{Y}^\alpha$
   - 2.4) Repeat Step 2.3) until $s(\underline{Y}^\alpha) - s(\underline{X}^\alpha) \leq \varepsilon$, then set $\underline{s}_N = s(\underline{Y}^\alpha)$.
3) Similarly, compute $\overline{s}_N = s(\overline{Y}^\alpha)$ based on $\overline{X}^\alpha = \left(F_j^{\leftarrow}(\alpha + \frac{(1-\alpha)i}{N}) \right)_{i,j}$.
4) Return $(\underline{s}_N, \; \overline{s}_N)$ (*rearrangement range*; taken as bounds on $\overline{\mathrm{VaR}}_\alpha(L)$)

- The RA aims at maximizing the minimal row sums (solving a *maximin problem*; minimax problem for $\underline{\mathrm{VaR}}_\alpha$).
- **Intuition:** A *completely mixable matrix* (equal row sums), would minimize the variance of $L \,|\, L > F_L^-(\alpha)$ and thus concentrate more of the $1-\alpha$ mass of F_L around the constant $\mathbb{E}[L \,|\, L > \mathrm{VaR}_\alpha(L)] \underset{\text{cont.}}{=} \mathrm{ES}_\alpha(L) \geq \mathrm{VaR}_\alpha(L)$, so $\mathrm{VaR}_\alpha(L)$ increases (F_L jumps to 1 in $\mathrm{VaR}_\alpha(L)$ so $\mathrm{VaR}_\alpha(L)$ is largest).



Example 8.25 (How the RA works)

1) Where it works (to compute the maximal minimal row sum):

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 2 \\ 3 & 5 & 4 \\ 4 & 7 & 8 \end{pmatrix} \Longrightarrow \sum_{-1} = \begin{pmatrix} 2 \\ 5 \\ 9 \\ 15 \end{pmatrix} \quad \begin{pmatrix} 4 & 1 & 1 \\ 3 & 3 & 2 \\ 2 & 5 & 4 \\ 1 & 7 & 8 \end{pmatrix} \overset{\text{here: stable}}{\Longrightarrow} \sum_{-2} = \begin{pmatrix} 5 \\ 5 \\ 6 \\ 9 \end{pmatrix} \quad \begin{pmatrix} 4 & 5 & 1 \\ 3 & 7 & 2 \\ 2 & 3 & 4 \\ 1 & 1 & 8 \end{pmatrix} \Longrightarrow \sum_{-3} = \begin{pmatrix} 9 \\ 10 \\ 5 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 5 & 2 \\ 3 & 7 & 1 \\ 2 & 3 & 4 \\ 1 & 1 & 8 \end{pmatrix} \overset{\text{here: not}}{\Longrightarrow} \sum_{-1} = \begin{pmatrix} 7 \\ 8 \\ 7 \\ 9 \end{pmatrix} \quad \begin{pmatrix} 3 & 5 & 2 \\ 2 & 7 & 1 \\ 4 & 3 & 4 \\ 1 & 1 & 8 \end{pmatrix} \quad \checkmark \Longrightarrow \sum = \begin{pmatrix} 10 \\ 10 \\ 11 \\ 10 \end{pmatrix} \quad \widehat{\overline{\mathrm{VaR}}}_\alpha(L^+) \approx 10$$

2) The RA can also fail:

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{pmatrix} \Longrightarrow \sum_{-1} = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} \quad \begin{pmatrix} 3 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 3 & 3 \end{pmatrix} \quad \checkmark \Longrightarrow \sum = \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} \quad \widehat{\overline{\mathrm{VaR}}}_\alpha(L^+) \approx 5 < 6$$

Example 8.26 ($\mathrm{Par}(\theta)$ margins)
Let $L_j \sim \mathrm{Par}(\theta)$ with $\bar{F}_j(x) = (1+x)^{-\theta}$, $j \in \{1,\dots,d\}$ (homogeneous case) and $\alpha = 0.999$. One obtains:

	$d=8$		$d=56$	
	$\theta=2$	$\theta=0.8$	$\theta=2$	$\theta=0.8$
$\overline{\mathrm{VaR}}_\alpha(L)$	465	300 182	3454	4 683 172
$\mathrm{VaR}_\alpha^+(L) = d\,\mathrm{VaR}_\alpha(L_1)$	245	44 979	1715	314 855
$\mathrm{VaR}_\alpha^\perp(L)$	96	75 877	293	862 855
$\underline{\mathrm{VaR}}_\alpha(L)$	31	5622	53	5622
$\overline{\mathrm{ES}}_\alpha(L) = d\,\mathrm{ES}_\alpha(L_1)$	498	–	3486	–
$\mathrm{ES}_\alpha^\perp(L)$	184	–	518	–
$\underline{\mathrm{ES}}_\alpha(L)$	178	–	472	–

- The "$+$" and "$\perp$" denote the comonotonic and independent case, resp.
- $\frac{\overline{\mathrm{ES}}_\alpha(L)}{\overline{\mathrm{VaR}}_\alpha(L)} \underset{d\uparrow\infty}{\approx} 1$ can be explained; see MFE (2015, Proposition 8.36).
- The dependence uncertainty spread $\overline{\mathrm{VaR}}_\alpha(L) - \underline{\mathrm{VaR}}_\alpha(L) \geq \overline{\mathrm{ES}}_\alpha(L) - \underline{\mathrm{ES}}_\alpha(L)$ can be explained; see MFE (2015, Proposition 8.37).

Remark 8.27

- The RA finds approximate solutions to *maximin* (for $\overline{\mathrm{VaR}}_\alpha(L)$) and *minimax* (for $\underline{\mathrm{VaR}}_\alpha(L)$) *problems* and is thus of wider interest (e.g., in Operations Research).
- For $\underline{\mathrm{ES}}_\alpha(L)$, discretize the whole support of each margin, rearrange, and approximate $\underline{\mathrm{ES}}_\alpha(L)$ by the nonparametric ES_α estimate of the row sums.
- The *Adaptive Rearrangement Algorithm (ARA)*
  - uses relative (instead of absolute) individual tolerances;
  - uses a relative joint tolerance to guarantee that $\underline{s}_N$ and $\overline{s}_N$ are close;
  - chooses N adaptively to reach the joint tolerance; and
  - determines convergence after each rearranged column.
- The *Block Rearrangement Algorithm* rearranges blocks of columns.

Example 8.28 (Superadditivity of VaR under special dependence)

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

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

from which one can show that

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

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

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

8.5 Capital allocation

How can the overall capital requirement may be disaggregated into additive contributions/units/investments? Motivation: How can we measure the risk-adjusted performance of different investments?

8.5.1 The allocation problem

- The performance of investments is usually measured using a *RORAC (return on risk-adjusted capital) approach* by considering

$$\frac{\text{expected profit of investment } j}{\text{risk capital for investment } j}.$$

- The risk capital of investment j with loss L_j can be computed as follows: Compute $\varrho(L) = \varrho(L_1 + \cdots + L_d)$. Then allocate $\varrho(L)$ to the investments according to a *capital allocation principle* such that

$$\varrho(L) = \sum_{j=1}^{d} \mathrm{AC}_j,$$

where the *risk contribution* AC_j is the capital allocated to investment j.

The formal set-up

- Consider an open set $\mathbf{1} \in \Lambda \subseteq \mathbb{R}^d \setminus \{\mathbf{0}\}$ of portfolio weights and define

$$L(\boldsymbol{\lambda}) = \boldsymbol{\lambda}' \boldsymbol{L} = \sum_{j=1}^{d} \lambda_j L_j, \quad \boldsymbol{\lambda} \in \Lambda.$$

- For a risk measure ϱ, define the *associated risk-measure function*

$$r_\varrho(\boldsymbol{\lambda}) = \varrho(L(\boldsymbol{\lambda})),$$

so that $r_\varrho(\mathbf{1}) = \varrho(L)$.

8.5.2 The Euler principle and examples

- If r_ϱ is positive homogeneous and differentiable at $\boldsymbol{\lambda} \in \Lambda$, Euler's rule implies that

$$r_\varrho(\boldsymbol{\lambda}) = \sum_{i=1}^{d} \lambda_i \frac{\partial r_\varrho}{\partial \lambda_i}(\boldsymbol{\lambda}) \quad \text{so} \quad \varrho(L) = r_\varrho(\mathbf{1}) = \sum_{j=1}^{d} \frac{\partial r_\varrho}{\partial \lambda_j}(\mathbf{1}).$$

Note that r_ϱ is positive homogeneous if ϱ is.

Definition 8.29 (Euler capital allocation principle)
If r_ϱ is a pos.-hom. risk-measure function and differentiable at $\boldsymbol{\lambda} = \mathbf{1}$, then the *Euler capital allocation principle* has risk contributions

$$\mathrm{AC}_j = \mathrm{AC}_j^\varrho := \frac{\partial r_\varrho}{\partial \lambda_j}(\mathbf{1}), \quad j \in \{1, \dots, d\}.$$

Example 8.30 (Covariance principle)
Consider $r_{\mathsf{SD}}(\boldsymbol{\lambda}) := \sqrt{\mathrm{var}(L(\boldsymbol{\lambda}))} = \sqrt{\boldsymbol{\lambda}' \Sigma \boldsymbol{\lambda}}$ where $\Sigma = \mathrm{cov}(\boldsymbol{L})$. Thus

$$\mathrm{AC}_j^\varrho = \frac{\partial r_{\mathsf{SD}}}{\partial \lambda_j}(\mathbf{1}) = \frac{(\Sigma \mathbf{1})_j}{r_{\mathsf{SD}}(\mathbf{1})} = \frac{\sum_{k=1}^d \mathrm{cov}(L_j, L_k)}{r_{\mathsf{SD}}(\mathbf{1})} = \frac{\mathrm{cov}(L_j, L)}{\sqrt{\mathrm{var}(L)}}.$$

Corollary 8.31 (Euler allocation under ellipticality)
Assume that r_ϱ is the risk-measure function of a positive-homogeneous and law invariant ϱ. Let $\boldsymbol{L} \sim E_d(\mathbf{0}, \Sigma, \psi)$. Then, under an Euler allocation,

$$\frac{\mathrm{AC}_j^\varrho}{\mathrm{AC}_k^\varrho} = \frac{\sum_{l=1}^d \Sigma_{jl}}{\sum_{l=1}^d \Sigma_{kl}}, \quad j, k \in \{1, \dots, d\}.$$

Example 8.32 (Euler allocation and shortfall contributions)
Now consider $r^{\alpha}_{\mathrm{ES}}(\boldsymbol{\lambda}) = \mathbb{E}(L \,|\, L \geq q_\alpha(L(\boldsymbol{\lambda})))$. Then

$$r^{\alpha}_{\mathsf{ES}}(\boldsymbol{\lambda}) = \frac{1}{1-\alpha}\int_\alpha^1 r^{u}_{\mathsf{VaR}}(\boldsymbol{\lambda})\,\mathrm{d}u,$$

Assuming the differentiability of $r^{u}_{\mathsf{VaR}}(\boldsymbol{\lambda})$, the Euler principle implies that

$$\frac{\partial r^{\alpha}_{\mathsf{ES}}}{\partial \lambda_j}(\mathbf{1}) = \frac{1}{1-\alpha}\int_\alpha^1 \frac{\partial r^{u}_{\mathsf{VaR}}}{\partial \lambda_j}(\mathbf{1})\,\mathrm{d}u = \frac{1}{1-\alpha}\int_\alpha^1 \mathbb{E}(L_j \,|\, L = F_L^{\leftarrow}(u))\,\mathrm{d}u.$$

If F_L has a differentiable inverse,

$$\frac{\partial r^{\alpha}_{\mathsf{ES}}}{\partial \lambda_j}(\mathbf{1}) = \frac{1}{1-\alpha}\int_{F_L^{\leftarrow}(\alpha)}^\infty \mathbb{E}(L_j \,|\, L = v) f_L(v)\,\mathrm{d}v = \frac{\mathbb{E}(L_j; L \geq F_L^{\leftarrow}(\alpha))}{1-\alpha}.$$

Hence the Euler capital allocation takes the form

$$\mathrm{AC}_j^{\varrho} = \mathbb{E}(L_j \,|\, L \geq \mathrm{VaR}_\alpha(L)), \quad L := L(\mathbf{1});$$

AC_j^{ϱ} is known as the *expected shortfall contribution* of investment j. This is a popular allocation principle in practice.

9 Market Risk

9.1 Risk factors and mapping

9.2 Market risk measurement

9.3 Backtesting

9.1 Risk factors and mapping

9.1.1 The loss operator

- The key idea in this section is that of a loss operator for expressing the change in value of a portfolio in terms of risk-factor changes.
- Let the current time be t and assume the current value V_t os an asset portfolio is known, or can be computed with appropriate valuation models.
- We are interested in value changes or losses over a relatively short time period $[t, t+1]$, for example one day, two weeks or month.
- Scaling may be applied to derive capital requirements for longer periods.
- We assume there is no change to the composition of the portfolio over the time period.
- The future value V_{t+1} is modelled as a random variable.

- We want to determine the distribution of the loss distribution of $L_{t+1} = -(V_{t+1} - V_t)$.
- We map the value at time t using the formula

$$V_t = g(\tau_t, \boldsymbol{Z}_t)$$

where τ_t is time t expressed in units of valuation time.

The issue of time

- We will be quite precise about the modelling of time.
- The natural time unit for valuation of positions might be yearly; e.g. in Black-Scholes valuation, the volatility is expressed in annualized terms.
- On the other hand the risk modelling time horizon $[t, t+1]$ is typically shorter.
- Let Δt be the length of the time horizon in valuation time.

- For example, suppose that valuation time is yearly. Then a monthly time horizon would be $\Delta t = 1/12$ and a trading day $\Delta t = 1/250$.
- We set $\tau_t = t(\Delta t)$ for all t so that $\tau_{t+1} - \tau_t = \Delta t$.

From the mapping to the loss operator

- The risk factor changes over the time horizon are

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

- Typically, historical risk factor data are available as a time series $\mathbf{X}_{t-n}, \ldots, \mathbf{X}_{t-1}, \mathbf{X}_t$ and these are used to model the behaviour of $\mathbf{X}_{t+1}$.
- We have

$$\begin{aligned} L_{t+1} &= -(V_{t+1} - V_t) \\ &= -\left(g(\tau_{t+1}, \mathbf{Z}_{t+1}) - g(\tau_t, \mathbf{Z}_t)\right) \\ &= -(g(\tau_t + \Delta t, \mathbf{Z}_t + \mathbf{X}_{t+1}) - g(\tau_t, \mathbf{Z}_t)). \end{aligned} \tag{40}$$

- Since the risk factor values $\mathbf{Z}_t$ are known at time t, the loss L_{t+1} is determined by the risk factor changes $\mathbf{X}_{t+1}$.
- Given a realization $\mathbf{z}_t$ of $\mathbf{Z}_t$, the loss operator at time t is defined to be

$$l_{[t]}(\mathbf{x}) = -(g(\tau_t + \Delta t, \mathbf{z}_t + \mathbf{x}) - g(\tau_t, \mathbf{z}_t)), \tag{41}$$

  so that

$$L_{t+1} = l_{[t]}(\mathbf{X}_{t+1}).$$

- The loss operator embodies the idea of full revaluation.
- From the perspective of time t the loss distribution of L_{t+1} is determined by the multivariate distribution of $\mathbf{X}_{t+1}$.

9.1.2 Delta and delta–gamma approximations

- If the mapping function g is differentiable and Δt is relatively small we can approximate g with a first-order Taylor series approximation

$$g(\tau_t + \Delta t, \boldsymbol{z}_t + \boldsymbol{x}) \approx g(\tau_t, \boldsymbol{z}_t) + g_\tau(\tau_t, \boldsymbol{z}_t)\Delta t + \sum_{i=1}^{d} g_{z_i}(\tau_t, \boldsymbol{z}_t)x_i, \tag{42}$$

where the τ-subscript and z_i-subscript denote partial derivatives with respect to (valuation) time and the risk factors respectively.

- This allows us to approximate the loss operator in (41) by the linear loss operator at time t given by

$$l^{\Delta}_{[t]}(\boldsymbol{x}) := -\left(g_\tau(\tau_t, \boldsymbol{z}_t)\Delta t + \sum_{i=1}^{d} g_{z_i}(\tau_t, \boldsymbol{z}_t) x_i \right). \tag{43}$$

- Note that, when working with a short time horizon Δt, the term $g_\tau(\tau_t, \boldsymbol{z}_t)\Delta t$ is typically small and is sometimes omitted in practice.

Example 9.1 (European call option)

- Consider portfolio consisting of one standard European call on a non-dividend paying stock S with maturity T and exercise price K.
- The Black-Scholes value of this asset at time t is $C^{BS}(t, S_t, r, \sigma)$ where

$$C^{BS}(t, S; r, \sigma) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2),$$

  Φ is standard normal df, r represents risk-free interest rate, σ the volatility of underlying stock, and where

$$d_1 = \frac{\log(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \text{ and } d_2 = d_1 - \sigma\sqrt{T-t}.$$

- While in the BS model, it is assumed that interest rates and volatilities are constant, in reality they tend to fluctuate over time; they should be added to our set of risk factors.

- The risk factors: $\mathbf{Z}_t = (\log S_t, r_t, \sigma_t)'$.
- The risk factor changes: $\mathbf{X}_t = (\log(S_t/S_{t-1}), r_t - r_{t-1}, \sigma_t - \sigma_{t-1})'$.
- The mapping:

$$V_t = C^{BS}(\tau_t, S_t; r_t, \sigma_t) = g(\tau_t, \boldsymbol{Z}_t)$$

- For derivative positions it is quite common to use the linear loss operator

$$L^{\Delta}_{t+1} = l^{\Delta}_{[t]}(\boldsymbol{X}_{t+1}) = -\left(g_\tau(\tau_t, \mathbf{z}_t)\Delta t + \sum_{i=1}^{3} g_{z_i}(\tau_t, \mathbf{z}_t) X_{t+1,i}\right),$$

  where g_τ, g_{z_i} denote partial derivatives.
- Δt is the length of the time interval expressed in years since Black-Scholes parameters relate to units of one year.

- It is more common to write the linear loss operator as

$$l^{\Delta}_{[t]}(\boldsymbol{x}) = -\left(C^{BS}_t + C^{BS}_S S_t x_1 + C^{BS}_r x_2 + C^{BS}_\sigma x_3\right),$$

in terms of the derivatives of the BS formula or the Greeks.

  - C^{BS}_S is known as the delta of the option.
  - C^{BS}_σ is the vega.
  - C^{BS}_r is the rho.
  - C^{BS}_t is the theta.

Note the appearance of S_t in the C^{BS}_S term. This is because the risk factor is $\ln S_t$ rather than S_t and $C^{BS}_{\ln S} = C^{BS}_S S_t$.

Quadratic loss operator

- Recall the first-order Taylor series approximation of mapping in (42).
- Let $\boldsymbol{\delta}(\tau_t, \boldsymbol{z}_t) = (g_{z_1}(\tau_t, \boldsymbol{z}_t), \ldots, g_{z_d}(\tau_t, \boldsymbol{z}_t))'$ be the first-order partial derivatives of the mapping with respect to the risk factors.

- Let $\boldsymbol{\omega}(\tau_t, \boldsymbol{z}_t) = (g_{z_1\tau}(\tau_t, \boldsymbol{z}_t), \ldots, g_{z_d\tau}(\tau_t, \boldsymbol{z}_t))'$ denote the mixed partial derivatives with respect to time and the risk factors.
- Let $\Gamma(\tau_t, \boldsymbol{z}_t)$ denote the matrix with (i, j)th element given by $g_{z_i z_j}(\tau_t, \boldsymbol{z}_t)$; this matrix contains gamma sensitivities to individual risk factors on the diagonal and cross gamma sensitivities to pairs of risk factors off the diagonal.
- The full second-order approximation of the mapping function is g is

$$\begin{aligned} g(\tau_t + \Delta t, \boldsymbol{z}_t + \boldsymbol{x}) \approx\ & g(\tau_t, \boldsymbol{z}_t) + g_\tau(\tau_t, \boldsymbol{z}_t)\Delta t + \boldsymbol{\delta}(\tau_t, \boldsymbol{z}_t)'\boldsymbol{x} + \\ & \frac{1}{2}(g_{\tau\tau}(\tau_t, \boldsymbol{z}_t)(\Delta t)^2 + 2\boldsymbol{\omega}(\tau_t, \boldsymbol{z}_t)'\boldsymbol{x}\Delta t + \\ & \boldsymbol{x}'\Gamma(\tau_t, \boldsymbol{z}_t)\boldsymbol{x}) \,. \end{aligned}$$

- In practice, we would usually omit terms of order $o(\Delta_t)$ (terms that tend to zero faster than Δ_t). In standard continuous-time financial models like Black-Scholes the risk-factor changes $\boldsymbol{x}$ are of order $\sqrt{\Delta_t}$.

- This leaves us with the quadratic loss operator

$$l_{[t]}^{\Delta\Gamma}(\boldsymbol{x}) = -(g_\tau(\tau_t, \boldsymbol{z}_t)\Delta t + \boldsymbol{\delta}(\tau_t, \boldsymbol{z}_t)'\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}'\Gamma(\tau_t, \boldsymbol{z}_t)\boldsymbol{x}) \tag{44}$$

  which is more accurate than the linear loss operator (43).

Example 9.2 (European call option)

The quadratic loss operator is

$$\begin{aligned} l_{[t]}^{\Delta\Gamma}(\boldsymbol{x}) =& l_{[t]}^{\Delta}(\boldsymbol{x}) - 0.5\left(C_{SS}^{BS}S_t^2x_1^2 + C_{rr}^{BS}x_2^2 + C_{\sigma\sigma}^{BS}x_3^2\right) \\ & - \left(C_{Sr}^{BS}S_tx_1x_2 + C_{S\sigma}^{BS}S_tx_1x_3 + C_{r\sigma}^{BS}x_2x_3\right). \end{aligned}$$

The names of the second-order Greeks (with the exception of gamma) are rather obscure. Here are some of them:

- C_{SS}^{BS} is known as the gamma of the option;
- $C_{\sigma\sigma}^{BS}$ is the vomma;
- $C_{S\sigma}^{BS}$ is the vanna.

9.1.3 Mapping bond portfolios

Basic definitions for bond pricing

- Let $p(t,T)$ denote the price at time t of a default-free zero-coupon bond paying one at time T (also called a discount factor).
- Time is measured in years.
- Many other fixed-income instruments such as coupon bonds or standard swaps can be viewed as portfolios of zero-coupon bonds.
- The mapping $T \to p(t,T)$ for different maturities is one way of describing the so-called term structure of interest rates at time t. An alternative description is based on yields.
- The term structure $T \to p(t,T)$ is known at time t.
- However the future term structure $T \to p(t+x,T)$ for $x > 0$ is not known at time t and must be modelled stochastically.

- The continuously compounded yield of a zero-coupon bond is

$$y(t,T) = -\frac{\ln p(t,T)}{T-t}. \tag{45}$$

- We have the relation

$$p(t,T) = \exp(-(T-t)y(t,T))\,.$$

- The yield is the constant, annualized rate implied by the price $p(t,T)$. Also known as spot rate.
- The mapping $T \to y(t,T)$ is referred to as the continuously compounded yield curve at time t.
- Yields are comparable across different times to maturity.

Detailed mapping of a bond portfolio

- Consider a portfolio of d default-free zero-coupon bonds with maturities T_i and prices $p(t,T_i)$ for $i = 1,\ldots,d$. Assume $p(T_i,T_i) = 1$ for all i.

- By λ_i we denote the number of bonds with maturity T_i in the portfolio.
- The portfolio value at time t is given by

$$V(t) := \sum_{i=1}^{d} \lambda_i p(t, T_i) = \sum_{i=1}^{d} \lambda_i \exp(-(T_i - t) y(t, T_i)).$$

- In a detailed analysis of the change in value one takes all yields $y(t, T_i)$, $1 \leq i \leq d$, as risk factors.
- We want to put this in the general discrete-time framework of the mapping

$$V_t = g(\tau_t, \boldsymbol{Z}_t).$$

- We set

$$\tau_t = t(\Delta t), \quad V_t = V(\tau_t), \quad Z_{t,i} = y(\tau_t, T_i)$$

where Δt is risk management time horizon in years.

- We obtain a mapping of the form

$$V_t = V(\tau_t) = g(\tau_t, \boldsymbol{Z}_t) = \sum_{i=1}^{d} \lambda_i \exp(-(T_i - \tau_t) Z_{t,i}). \tag{46}$$

The loss operator and its approximations

- The portfolio loss is

$$\begin{aligned} L_{t+1} &= -(V_{t+1} - V_t) \\ &= -\sum_{i=1}^{d} \lambda_i e^{-(T_i - \tau_t) Z_{t,i}} \Big(\exp\left(Z_{t,i} \Delta t - (T_i - \tau_{t+1}) X_{t+1,i}\right) - 1 \Big) . \end{aligned}$$

- Reverting to standard bond pricing notation the loss operator is

$$l_{[t]}(\boldsymbol{x}) = -\sum_{i=1}^{d} \lambda_i p(\tau_t, T_i) \Big(\exp\left(y(\tau_t, T_i) \Delta t - (T_i - \tau_{t+1}) x_i\right) - 1 \Big) ,$$

where x_i represents the change in yield of the ith bond.

- The first derivatives of the mapping function (46) are

$$g_\tau(\tau_t, \boldsymbol{z}_t) = \sum_{i=1}^{d} \lambda_i p(\tau_t, T_i) z_{t,i}$$
$$g_{z_i}(\tau_t, \boldsymbol{z}_t) = -\lambda_i (T_i - \tau_t) \exp(-(T_i - \tau_t) z_{t,i}).$$

- Inserting these in (43) and reverting to standard bond pricing notation we obtain

$$l^{\Delta}_{[t]}(\boldsymbol{x}) = -\sum_{i=1}^{d} \lambda_i p(\tau_t, T_i) \Big(y(\tau_t, T_i) \Delta t - (T_i - \tau_t) x_i \Big), \qquad (47)$$

- For the second-order approximation we need the second derivatives with respect to yields which are

$$g_{z_i z_i}(\tau_t, \boldsymbol{z}_t) = \lambda_i (T_i - \tau_t)^2 \exp(-(T_i - \tau_t) z_{t,i})$$

and $g_{z_i z_j}(\tau_t, \boldsymbol{z}_t) = 0$ for $i \neq j$.

- The quadratic loss operator (44) is

$$l_{[t]}^{\Delta\Gamma}(\boldsymbol{x}) = -\sum_{i=1}^{d} \lambda_i p(\tau_t, T_i) \left(y(\tau_t, T_i)\Delta t - (T_i - \tau_t)x_i + \frac{1}{2}(T_i - \tau_t)^2 x_i^2 \right). \tag{48}$$

Relationship of linear operator to duration

- Consider a very simple model for the yield curve at time t in which

$$y(\tau_{t+1}, T_i) = y(\tau_t, T_i) + x$$

  for all maturities T_i.

- In our mapping notation

$$Z_{t+1,i} = Z_{t,i} + X_{t+1}, \quad \forall i.$$

- In this model we assume that a parallel shift in level takes place along the entire yield curve.

- This is unrealistic but frequently assumed in practice.
- In this model the loss operator and its linear and quadratic approximations are functions of a scalar variable x, the change in level.
- Under the parallel shift model we can write

$$l^{\Delta}_{[t]}(x) = -V_t\Big(A_t\Delta t - D_t x\Big), \tag{49}$$

where

$$D_t := \sum_{i=1}^{d} \frac{\lambda_i p(\tau_t, T_i)}{v_t}(T_i - \tau_t), \quad A_t := \sum_{i=1}^{d} \frac{\lambda_i p(\tau_t, T_i)}{V_t} y(\tau_t, T_i).$$

- D_t is usually called the (Macaulay) duration of the bond portfolio.
- It is a weighted sum of the times to maturity of the different cash flows in the portfolio, the weights being proportional to the discounted values of the cash flows.

9.2 Market risk measurement

The goal in this section is to estimate the distribution of

$$L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$$

or a linear or quadratic approximation thereof, where

- $\boldsymbol{X}_{t+1}$ is the vector of risk-factor changes from time t to time $t+1$;
- $l_{[t]}$ is the known loss operator function at time t.

The problem comprises two tasks:

1) on the one hand we have the statistical problem of estimating the distribution of $\boldsymbol{X}_{t+1}$;
2) on the other hand we have the computational or numerical problem of evaluating the distribution of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$.

9.2.1 Conditional and unconditional loss distributions

- Generally, we want to compute conditional measures of risk based on the most recent information about financial markets.
- In this case, the task is to estimate $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$, the conditional distribution of risk-factor changes, given $\mathcal{F}_t$, the sigma field representing the available information at time t.
- The conditional loss distribution is the distribution of the loss operator $l_{[t]}(\cdot)$ under $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$, i.e. the distribution with df

$$F_{L_{t+1}|\mathcal{F}_t}(l) = \mathbb{P}(l_{[t]}(\boldsymbol{X}_{t+1}) \leq l \,|\, \mathcal{F}_t).$$

- In the unconditional approach we assume that $(\boldsymbol{X}_s)_{s\leq t}$ forms a stationary time series, at least in the recent past.
- In this case we can estimate the stationary distribution $F_{\boldsymbol{X}}$ and then evaluate the unconditional loss distribution of $l_{[t]}(\boldsymbol{X})$ where $\boldsymbol{X} \sim F_{\boldsymbol{X}}$. The unconditional loss distribution is thus $F_{L_{t+1}}(l) = \mathbb{P}(l_{[t]}(\boldsymbol{X}) \leq l)$.

- The unconditional approach may be appropriate for longer time intervals, or for stress testing during quieter periods.
- If the risk-factor changes form an iid series, we obviously have $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t} = F_{\boldsymbol{X}}$, so that the conditional and unconditional approaches coincide.

9.2.2 Variance-covariance method

- The variance–covariance method is an analytical method in which strong assumptions of (conditional) normality and linearity are made.
- We assume that the conditional distribution of risk-factor changes $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$ is a multivariate normal distribution.
- In other words, we assume that $\boldsymbol{X}_{t+1} \,|\, \mathcal{F}_t \sim \mathrm{N}_d(\boldsymbol{\mu}_{t+1}, \Sigma_{t+1})$.
- The estimation of $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$ can be carried out in a number of ways:
  - Fit multivariate ARMA-GARCH model with multivariate normal innovations; use model to derive estimates of $\boldsymbol{\mu}_{t+1}$ and Σ_{t+1}.

- Alternatively use the exponentially weighted moving-average (EWMA) procedure; Σ_{t+1} estimated recursively by $\hat{\Sigma}_{t+1} = \theta \boldsymbol{X}_t \boldsymbol{X}_t' + (1-\theta)\hat{\Sigma}_t$ where θ is a small positive number (typically $\theta \approx 0.04$).

- The second critical assumption in the variance–covariance method is that the linear loss operator is sufficiently accurate. The linear loss operator is a function of the form

$$l^{\Delta}_{[t]}(\boldsymbol{x}) = -(c_t + \boldsymbol{b}_t' \boldsymbol{x})$$

for some constant c_t and constant vector $\boldsymbol{b}_t$, known at time t.

- We infer that, conditional on $\mathcal{F}_t$,

$$L^{\Delta}_{t+1} = l^{\Delta}_{[t]}(\boldsymbol{X}_{t+1}) \sim \mathrm{N}(-c_t - \boldsymbol{b}_t' \boldsymbol{\mu}_{t+1},\ \boldsymbol{b}_t' \Sigma_{t+1} \boldsymbol{b}_t).$$

- Under normality, VaR_α and ES_α may be easily calculated:
  - $\widehat{\mathrm{VaR}}_\alpha = -c_t - \boldsymbol{b}_t' \widehat{\boldsymbol{\mu}}_{t+1} + \sqrt{\boldsymbol{b}_t' \widehat{\Sigma}_{t+1} \boldsymbol{b}_t}\, \Phi^{-1}(\alpha)$.
  - $\widehat{\mathrm{ES}}_\alpha = -c_t - \boldsymbol{b}_t' \widehat{\boldsymbol{\mu}}_{t+1} + \sqrt{\boldsymbol{b}_t' \widehat{\Sigma}_{t+1} \boldsymbol{b}_t}\, \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}$.

Pros and cons, extensions

Pros: In contrast to the methods that follow, variance-covariance offers analytical solution with no simulation.

Cons:

- Assumption of multivariate normality may seriously underestimate the tail of the loss distribution.
- Linearization may be a crude approximation.

Extensions: Instead of assuming normal risk factors, the method could be easily adapted to use multivariate Student t or multivariate hyperbolic risk-factor changes without sacrificing tractability (the method works for all elliptical distributions but linearization is crucial here).

9.2.3 Historical simulation

- Historical simulation is by far the most popular method used by banks for the trading book.
- Instead of estimating the distribution of $l_{[t]}(\boldsymbol{X}_{t+1})$ under an explicit parametric model for $\boldsymbol{X}_{t+1}$, the historical simulation method can be thought of as estimating the distribution of the loss operator under the *empirical distribution* of historical data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$.
- Construct the historically simulated losses (under the current portfolio):

$$\{\tilde{L}_s = l_{[t]}(\boldsymbol{X}_s) : s = t-n+1, \ldots, t\}.$$

- One may apply the linear/quadratic loss operator (if that was already used; avoids revaluation).
- $\tilde{L}_s$ shows what would happen to the current portfolio if the risk-factor change on day s were to recur.
- Use $(\tilde{L}_s)$ to make inferences about the loss distribution and risk measures.

- **Inference about the loss distribution**
  - One could use empirical quantile estimation to estimate VaR_α.
    But: What about precision (sample size; confidence intervals)?
  - Or fit a parametric distribution to the historical losses $L_{t-n+1}, \ldots, L_t$ and calculate risk measures from this distribution.
    But: Which distribution to fit (body or tail)?
  - One could use extreme value theory to estimate the tail of the loss distribution and related risk measures based on the historical losses $L_{t-n+1}, \ldots, L_t$.

Theoretical justification

If $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$ are iid or, more generally, stationary, convergence of the empirical distribution to the true distribution is ensured by a suitable version of the Law of Large Numbers (e.g. Glivenko–Cantelli theorem).

Pros and Cons

Pros:
- Easy to implement.
- No statistical estimation of the distribution of $\boldsymbol{X}$ necessary (the empirical df of $\boldsymbol{X}$ is used implicitly).

Cons:
- It may be difficult to collect sufficient quantities of relevant, synchronized data for all risk factors.
- Historical data may not contain examples of extreme scenarios ("driving a car by only looking in the back mirror").

Note:
- The dependence here is given by the empirical df of $\boldsymbol{X}$.
- "Historical simulation method" is a bit of a misnomer; there is no simulation in the sense of random number generation.

9.2.4 Dynamic Historical Simulation

- Assume that $\{\tilde{L}_s = l_{[t]}(\boldsymbol{X}_s) : s = t-n+1, \ldots, t\}$ are realizations from a stationary process $(\tilde{L}_s)$ of the form $\tilde{L}_s = \mu_s + \sigma_s Z_s$, where
  - μ_s and σ_s are $\mathcal{F}_{s-1}$-measurable;
  - (Z_s) are $\mathrm{SWN}(0,1)$ innovations with distribution function F_Z.

  Example: ARMA-GARCH model.
- We can easily calculate that for the next loss $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$ ahead

$$F_{L_{t+1}|\mathcal{F}_t}(l) = \mathbb{P}(\mu_{t+1} + \sigma_{t+1} Z_{t+1} \le l \mid \mathcal{F}_t) = F_Z((l - \mu_{t+1})/\sigma_{t+1}).$$

- Writing VaR_α^t for $F^{\leftarrow}_{L_{t+1}|\mathcal{F}_t}(\alpha)$ and ES_α^t for ES, we obtain

$$\mathrm{VaR}_\alpha^t = \mu_{t+1} + \sigma_{t+1}\,\mathrm{VaR}_\alpha(Z),$$
$$\mathrm{ES}_\alpha^t = \mu_{t+1} + \sigma_{t+1}\,\mathrm{ES}_\alpha(Z),$$

  where Z is a random variable with distribution function F_Z.

- **Estimation**
  - Formal parametric time series modelling to estimate μ_{t+1}, σ_{t+1}, $\mathrm{VaR}_\alpha(Z)$ and $\mathrm{ES}_\alpha(Z)$.
  - Often $\mu_{t+1} \approx 0$ and can be neglected. We can use EWMA to estimate $\sigma_{t-n+1}, \ldots, \sigma_t, \sigma_{t+1}$ and use the standardized residuals $\{\widehat{Z}_s = \tilde{L}_s/\hat{\sigma}_s,\ s = t-n+1, \ldots, t\}$ to estimate $\mathrm{VaR}_\alpha(Z)$ and $\mathrm{ES}_\alpha(Z)$.

9.2.5 Monte Carlo

- Estimate the distribution of $L = \ell_{[t]}(\boldsymbol{X}_{t+1})$ under some explicit parametric model for $\boldsymbol{X}_{t+1}$.
- In contrast to the variance-covariance approach we do not necessarily make the problem analytically tractable by linearizing the loss and making an assumption of normality for the risk factors.
- Instead, make inference about L using simulated risk factor data.

The method

1) Based on the historical risk-factor data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$, estimate a suitable statistical model for the risk-factor changes.
2) Simulate N new risk-factor changes $\boldsymbol{X}_{t+1}^{(1)}, \ldots, \boldsymbol{X}_{t+1}^{(N)}$ from this model.
3) Construct the simulated losses $L_k = \ell_{[t]}(\boldsymbol{X}_{t+1}^{(k)})$, $k \in \{1, \ldots, N\}$.
4) Make inference about the loss distribution F_L and risk measures using L_k, $k \in \{1, \ldots, N\}$ (similar possibilities as for the historical simulation method: non-parametric/parametric/EVT).

Pros and Cons

Pros:
- General. Any distribution for $\boldsymbol{X}_{t+1}$ can be taken.

Cons:
- Can be time consuming if loss operator is difficult to evaluate (depends on size and complexity of the portfolio).
- Note that MC approach does not address the problem of determining the distribution of $\boldsymbol{X}_{t+1}$.

9.3 Backtesting

- Backtesting is the practice of evaluating risk measurement procedures by comparing ex ante estimates/forecasts of risk measures with ex post realized losses and gains.
- It allows us to evaluate whether a model and estimation procedure produce credible risk measure estimates.

9.3.1 Violation-based tests for VaR

- Let VaR_α^t denote the α-quantile of the conditional loss distribution $F_{L_{t+1}|\mathcal{F}_t}$ and consider the event indicator variable $I_{t+1} = I_{\{L_{t+1} > \mathrm{VaR}_\alpha^t\}}$.
- The event $\{L_{t+1} > \mathrm{VaR}_\alpha^t\}$ is a VaR violation or exception.
- Assuming a continuous loss distribution, we have, by definition of the quantile,

$$\mathbb{E}(I_{t+1} \mid \mathcal{F}_t) = \mathbb{P}(L_{t+1} > \mathrm{VaR}_\alpha^t \mid \mathcal{F}_t) = 1 - \alpha\,, \tag{50}$$

- I_{t+1} is a Bernoulli variable with event probability $(1-\alpha)$.
- Moreover, the sequence of VaR exception indicators (I_t) is an iid sequence.
- The sum of exception indicators is binomially distributed:

$$M = \sum_{t=1}^{m} I_t \sim B(m, 1-\alpha).$$

- Assume exceptions occur at times $1 \leq T_1 < \cdots < T_M \leq m$ and set $T_0 = 0$. The spacings $S_j = T_j - T_{j-1}$ will be independent geometrically distributed rvs with mean $1/(1-\alpha)$, so that

$$\mathbb{P}(S_j = k) = \alpha^{k-1}(1-\alpha), \quad k \in \mathrm{N}\,.$$

- Both of these properties are testable in empirical data.
- For small event probability $1-\alpha$, the Bernoulli Trials Process may be well approximated by a Poisson process.

- Also for small $1-\alpha$ the geometric distribution may be approximated by an exponential distribution.
- Suppose we estimate VaR_α^t at time point t by $\widehat{\mathrm{VaR}}_\alpha^t$.
- In a backtest we consider empirical indicator variables

$$\widehat{I}_{t+1} = I_{\{L_{t+1} > \widehat{\mathrm{VaR}}_\alpha^t\}}.$$

- The sequence $(\widehat{I}_t)_{1 \le t \le m}$ should behave like a realization from a Bernoulli trials process with event probability $(1-\alpha)$.
- To test binomial behaviour for number of violations we compute a score test statistic

$$Z_m = \frac{(\sum_{t=1}^m \hat{I}_t) - m(1-\alpha)}{\sqrt{m\alpha(1-\alpha)}}$$

and reject Bernoulli hypothesis at 5% level if $Z_m > \Phi^{-1}(0.95)$.

- Exponential spacings can be tested numerically or with a Q-Q plot.

9.3.2 Violation-based tests of expected shortfall

- Let ES_α^t denote the one-period expected shortfall and $\widehat{\mathrm{ES}}_\alpha^t$ its estimate.
- Assume (L_t) follows a model of the form $L_t = \sigma_t Z_t$, where σ_t is a function of $\mathcal{F}_{t-1}$ and the (Z_t) are $\mathrm{SWN}(0,1)$ innovations.
- Then we can define a process (K_t) by

$$K_{t+1} = \frac{(L_{t+1} - \mathrm{ES}_\alpha^t)}{\mathrm{ES}_\alpha^t} I_{\{L_{t+1} > \mathrm{VaR}_\alpha^t\}} = \frac{Z_{t+1} - \mathrm{ES}_\alpha(Z)}{\mathrm{ES}_\alpha(Z)} I_{\{Z_{t+1} > q_\alpha(Z)\}},$$

  and note that it is a zero-mean iid sequence.
- This suggests we form violation residuals of the form

$$\widehat{K}_{t+1} = \frac{(L_{t+1} - \widehat{\mathrm{ES}}_\alpha^t)}{\widehat{\mathrm{ES}}_\alpha^t} \widehat{I}_{t+1}. \tag{51}$$

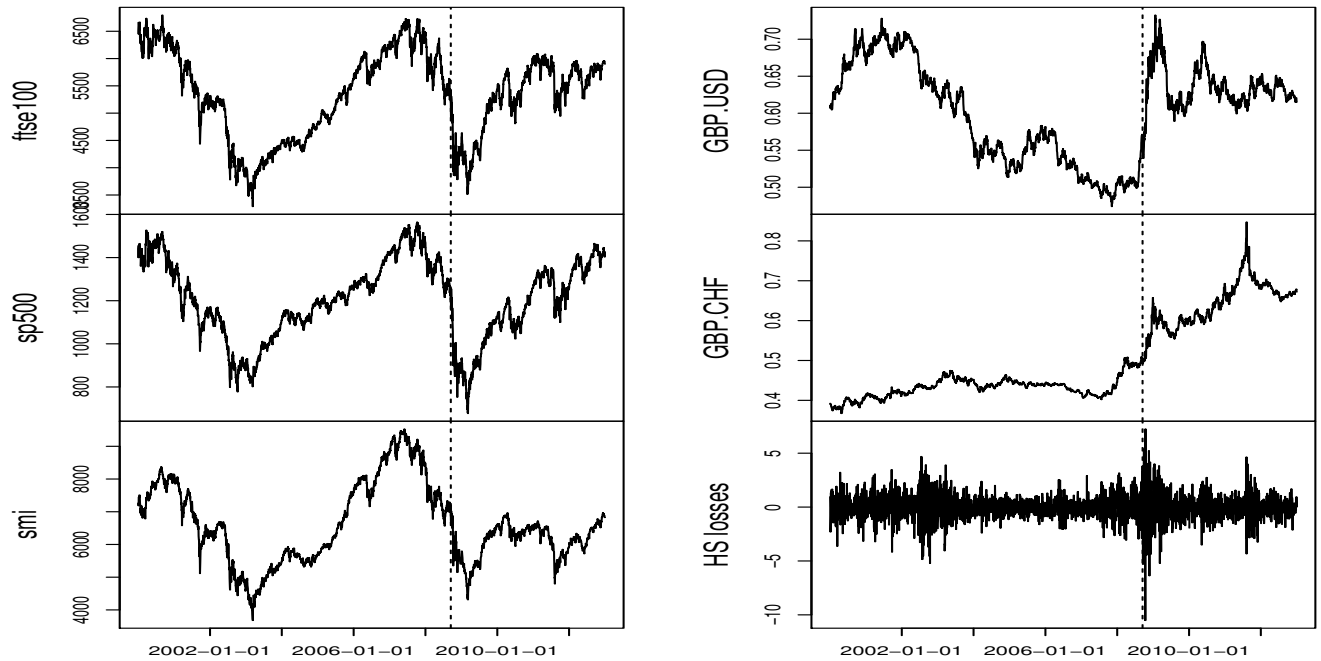
- We test for mean-zero behaviour using a bootstrap test on the non-zero violation residuals (McNeil and Frey (2000)).

9.3.3 Empirical comparison of methods using backtesting concepts

- We apply various VaR estimation methods to the portfolio of a hypothetical investor in international equity indexes.
- The investor is assumed to have domestic currency sterling (GBP) and to invest in the Financial Times 100 Shares Index (FTSE 100), the Standard & Poor's 500 (S&P 500) and the Swiss Market Index (SMI).
- The portfolio is influenced by five risk factors.
- On any day t we standardize the total portfolio value V_t in sterling to be one and assume portfolio weights are 30%, 40% and 30%, respectively.
- The loss operator and linear loss operator are:

$$l_{[t]}(\boldsymbol{x}) = 1 - (0.3e^{x_1} + 0.4e^{x_2+x_4} + 0.3e^{x_3+x_5})$$
$$l^{\Delta}_{[t]}(\boldsymbol{x}) = -(0.3x_1 + 0.4(x_2 + x_4) + 0.3(x_3 + x_5))$$

- x_1, x_2 and x_3 represent log-returns on the three indexes and x_4 and x_5 are log-returns on the GBP/USD and GBP/CHF exchange rates.



- The final picture shows the corresponding historical simulation data. The vertical dashed line is Lehman Brothers bankruptcy.

Estimation methods:

VC. The variance–covariance method assuming multivariate Gaussian risk-factor changes and using the multivariate EWMA method to estimate the conditional covariance matrix of risk-factor changes.

HS. The standard unconditional historical simulation method.

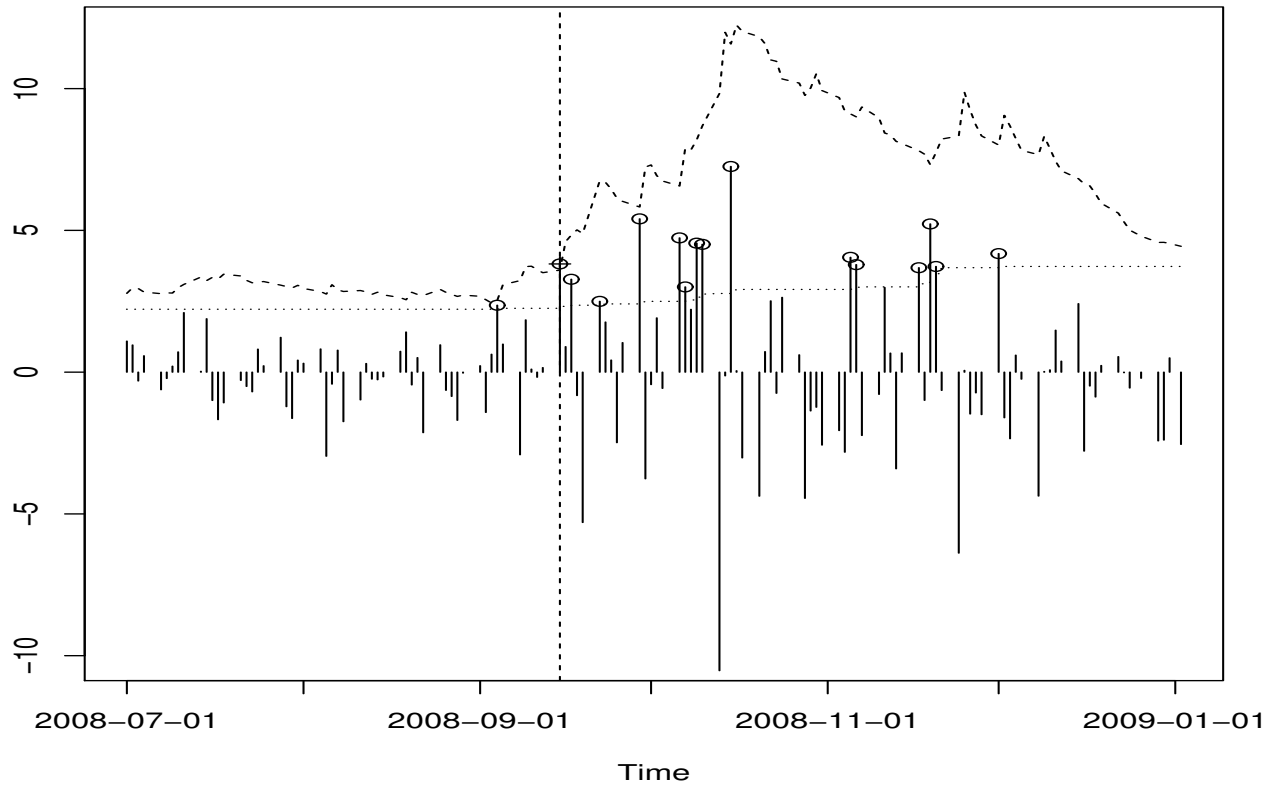
HS-GARCH. The univariate dynamic approach to historical simulation in which a $\mathrm{GARCH}(1,1)$ model with a constant conditional mean term and Gaussian innovations is fitted to the historically simulated losses to estimate the volatility of the next day's loss.

HS-GARCH-t. A similar method to HS-GARCH but Student t innovations are assumed in the GARCH model.

HS-MGARCH. The multivariate dynamic approach to historical simulation in which $\mathrm{GARCH}(1,1)$ models with constant conditional mean terms are fitted to each time series of risk-factor changes to estimate volatilities.

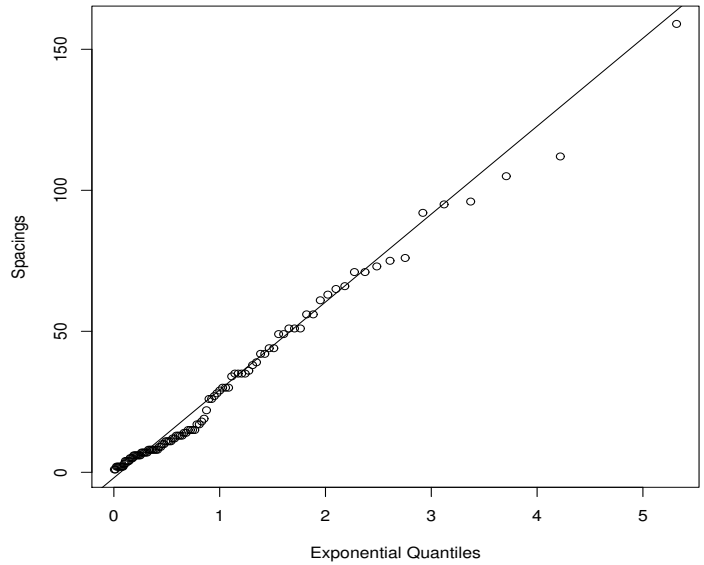
Year	2005	2006	2007	2008	2009	2010	2011	2012	All
Trading days	258	257	258	259	258	259	258	258	2065
Results for 95% VaR									
Expected no. of violations	13	13	13	13	13	13	13	13	103
VC	8	16	17	19	13	15	14	14	116
HS	0	6	28	49	19	6	10	1	119
HS-GARCH	9	13	22	22	13	14	9	15	117
HS-GARCH-t	9	14	23	22	14	15	10	15	122
HS-MGARCH	5	14	21	19	12	9	11	12	103
Results for 99% VaR									
Expected no. of violations	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	21
VC	2	8	8	8	2	4	5	6	43
HS	0	0	10	22	2	0	2	0	36
HS-GARCH	2	8	8	10	5	4	3	3	43
HS-GARCH-t	2	8	6	8	1	4	2	1	32
HS-MGARCH	0	4	4	5	0	1	2	1	17

- The HS method does not react to changing volatility:



- Dotted line is HS; dashed line is HS-MGARCH; vertical line is Lehmann.
- Circle is VaR violation for HS; cross is VaR violation for HS-MGARCH.

- Q-Q plot of spacings between exceptions (for HS-MGARCH):



- Violation residual test for ES (n: number of VaR violations):

	$\mathrm{ES}_{0.95}$	n	$\mathrm{ES}_{0.99}$	n
VC	0.00	116	0.05	43
HS	0.02	119	0.25	36
HS-GARCH	0.00	117	0.05	43
HS-GARCH-t	0.12	122	0.68	32
HS-MGARCH	0.99	103	0.55	17

10 Credit risk

10.1 Credit risky instruments

10.2 Measuring credit quality

10.3 Structural models of default

10.4 Bond and CDS pricing in hazard rate models

What is credit risk?

> "*Credit risk is the risk of a loss arising from the failure of a counterparty to honour its contractual obligations. This subsumes both default risk (the risk of losses due to the default of a borrower or a trading partner) and downgrade risk (the risk of losses caused by a deterioration in the credit quality of a counterparty that translates into a downgrading in some rating system).* "

- *Obligor* = a counterparty who has a financial obligation to us; for example, a debtor who owes us money, a bond issuer who promises interest, or a counterparty in a derivatives transaction.
- *Default* = failure to fulfill that obligation, for example, failure to repay loan or pay interest/coupon on a loan/bond; generally due to lack of liquidity or insolvency; may entail bankruptcy.

A crucial risk category

- A portfolio of loans or (corporate) bonds is obviously affected by credit risk.
- Credit risk accompanies any OTC (over-the-counter) derivative transaction such as a swap, because the default of one of the parties involved may substantially affect the actual pay-off of the transaction.
- There is a specialized market for *credit derivatives*, such as credit default swaps.
- Credit risk relates to the core activities of most banks but is also highly relevant to insurance companies: Insurers are exposed to substantial credit risk in their investment portfolios and counterparty default risk in their reinsurance treaties.

Credit risk management: A range of tasks

- An enterprise needs to determine the capital it requires to absorb losses due to credit risk.
- Portfolios of credit-risky instruments should be well diversified and optimized according to risk-return considerations.
- Institutions need to manage their portfolio of traded credit derivatives, which involves pricing, hedging and managing collateral for such trades.
- Financial institutions need to control the counterparty credit risk in their trades and contracts with other institutions. This has particularly been the case since the 2007–2009 financial crisis.

10.1 Credit risky instruments

This comprises loans, bonds, derivatives subject to counterparty risk and credit derivatives such as CDSs.

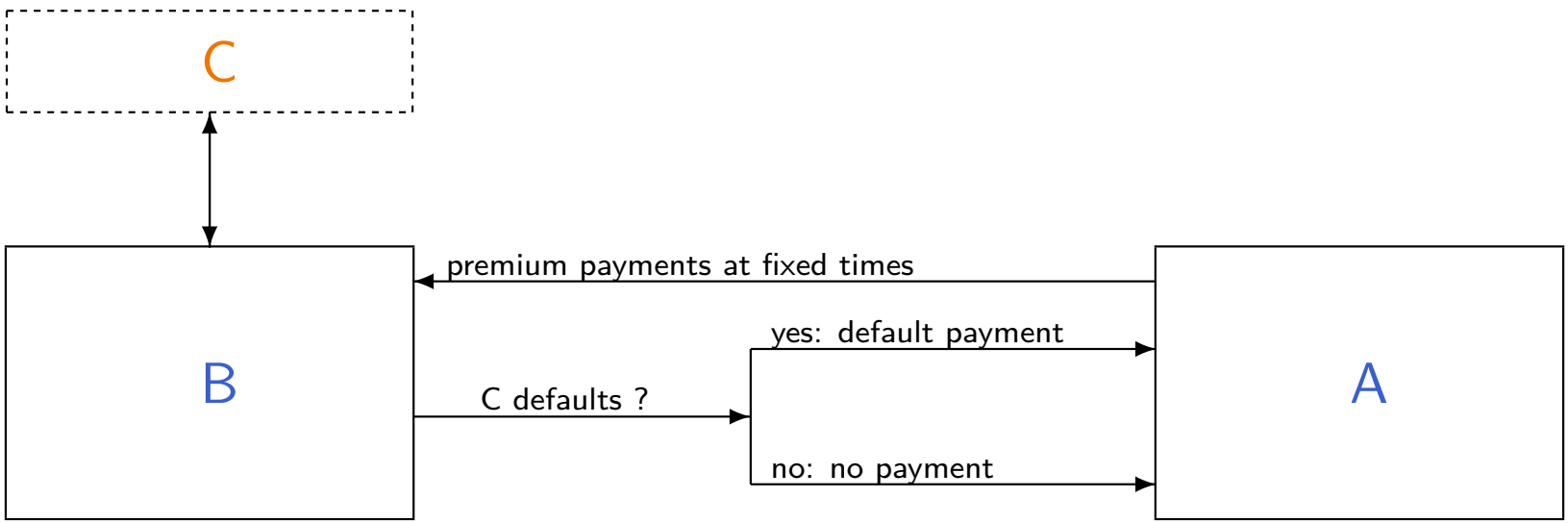
10.1.1 Credit default swaps and related credit derivatives

- Credit derivatives are securities which are primarily used for the hedging and trading of credit risk.
- The promised pay-off of a credit derivative is related to credit events affecting one or more firms.
- Major participants in the market for credit derivatives are banks, insurance companies and investment funds.
- Retail banks are typically net buyers of protection against credit events; other investors such as hedge funds and investment banks often act as both sellers and buyers of credit protection.
- Credit default swaps (CDSs) are the workhorses of the credit derivatives market and the market for CDSs written on larger corporations is fairly liquid.

Structure of CDS

Consider contract with maturity T and ignore counterparty credit risk. Three parties are involved (only two directly):

- C (reference entity); default at time $\tau_C < T$ triggers default payment.
- A (protection buyer); pays premiums to B until $\min(\tau_C, T)$.
- B (protection seller); makes default payment to A if $\tau_C < T$.



CDS: Payment flows

- If the reference entity experiences a default before the maturity date T of the contract, the protection seller makes a default payment to the protection buyer, which mimics the loss due to the default of a bond issued by the reference entity (the reference asset); this part of a CDS is called the default payment leg.
- As compensation the protection buyer makes periodic premium payments (typically quarterly or semiannually) to the protection seller (the premium payment leg); after the default of the reference entity, premium payments stop. There is no initial payment.
- The premium payments are quoted in the form of an annualized percentage x^* of the notional value of the reference asset; x^* is termed the (fair or market quoted) CDS spread.

Use of CDS

Investors enter into CDS contracts for various reasons.

- Bond investors with a large credit exposure to the reference entity may buy CDS protection to insure themselves against losses due to default of a bond (easier than reducing the original bond position as CDS contracts are more liquid).
- CDS contracts are also held for speculative reasons: so-called naked CDS positions, where the protection buyer does not own the bond are often assumed by investors who are speculating on the widening of the credit spread of the reference entity (similar to short-selling bonds issued by the reference entity.)
- Note that, in contrast to insurance, there is no requirement for the protection buyer to have insurable interest, that is, to actually own a bond issued by the reference entity.

10.2 Measuring credit quality

Scores, ratings & measures inferred from prices

There are two philosophies for quantifying the credit quality or default probability of an obligor.

1) Credit quality can be described by a credit *rating or score* that is based on empirical data and expert judgement.
2) For obligors whose equity is traded on financial markets, prices can be used to infer the market's view of the credit quality of the obligor.

Credit ratings and scores fulfill a similar function—they allow us to order obligors by their credit risk and map that risk to an estimate of the PD.

Credit ratings tend to be expressed on an ordered categorical scale whereas credit scores are often expressed in points on a metric scale.

Rating and scoring

- The task of rating obligors is often outsourced to a rating agency such as Moody's or Standard & Poor's (S&P).
- In the S&P rating system there are seven pre-default rating categories labelled AAA, AA, A, BBB, BB, B, CCC, with AAA being the highest and CCC the lowest rating.
- Moody's uses nine pre-default rating categories labelled Aaa, Aa, A, Baa, Ba, B, Caa, Ca, C.
- A finer alpha-numeric system is also used by both agencies.
- Credit scores are traditionally used for retail customers and are based on so-called *scorecards*. Historical data is used to model default risk as a function of demographic, behavioural and financial covariates using techniques like logistic regression. The covariates are weighted and combined into a score.

10.2.1 Credit rating migration

- In the credit-migration approach each firm is assigned to a credit-rating category at any given time point.
- We assume that the current credit rating completely determines the default probability.
- The probability of moving from one credit rating to another over a given risk horizon (typically one year) is then specified.
- These probabilities, known as transition probabilities, are typically presented in the form of a matrix. They are estimated from historical data on empirical transition rates.
- The following example is taken from Ou (2013), (Exhibit 26). It gives average transition rates from one rating to another within one year. WR stands for withdrawn rating.

Initial rating	Rating at year-end (%)									WR
	Aaa	Aa	A	Baa	Ba	B	Caa	Ca–C	Default	
Aaa	87.20	8.20	0.63	0.00	0.03	0.00	0.00	0.00	0.00	3.93
Aa	0.91	84.57	8.43	0.49	0.06	0.02	0.01	0.00	0.02	5.48
A	0.06	2.48	86.07	5.47	0.57	0.11	0.03	0.00	0.06	5.13
Baa	0.039	0.17	4.11	84.84	4.05	7.55	1.63	0.02	0.17	5.65
Ba	0.01	0.05	0.35	5.52	75.75	7.22	0.58	0.07	1.06	9.39
B	0.01	0.03	0.11	0.32	4.58	73.53	5.81	0.59	3.85	11.16
Caa	0.01	0.02	0.02	0.12	0.38	8.70	61.71	3.72	13.34	12.00
Ca-C	0.00	0.00	0.00	0.00	0.40	2.03	9.38	35.46	37.93	14.80

- 1-year default probability for an A-rated company is estimated to be 0.06%, whereas for a Caa-rated company it is 13.3%.
- In practice a correction to the figures would probably be undertaken to account for rating withdrawals

- Rating agencies also publish cumulative default rates over longer time horizons.
- These provide estimates of cumulative default probabilities over several years. Alternative estimates of multi-year default probabilities can be inferred from one-year transition matrices as explained later.
- The data are taken from Ou (2013), (Exhibit 33).

Initial rating	Term						
	1	2	3	4	5	10	15
Aaa	0.00	0.01	0.01	0.04	0.11	0.50	0.93
Aa	0.02	0.07	0.14	0.26	0.38	0.92	1.75
A	0.06	0.20	0.41	0.63	0.87	2.48	4.26
Baa	0.18	0.50	0.89	1.37	1.88	4.70	8.62
Ba	1.11	3.08	5.42	7.93	10.18	19.70	29.17
B	4.05	9.60	15.22	20.13	24.61	41.94	52.22
Caa-C	16.45	27.87	36.91	44.13	50.37	69.48	79.18

10.2.2 Rating transitions as a Markov chain

- Let (R_t) denote a discrete-time stochastic process taking values in $S = \{0, 1, \ldots, n\}$ at times $t = 0, 1, \ldots$.
- The set S defines rating states of increasing creditworthiness with 0 representing default. (R_t) models an obligor's rating over time.
- We will assume that (R_t) is a Markov chain. This means that it has the Markov property that

$$\mathbb{P}(R_t = k \mid R_0 = r_0, R_1 = r_1, \ldots, R_{t-1} = j) = \mathbb{P}(R_t = k \mid R_{t-1} = j)$$

  for all $t \geq 1$ and all $j, r_0, r_1, r_{t-2}, k \in S$.
- Conditional probabilities of rating transitions given an obligors's rating history depend only on the previous rating $R_{t-1} = j$ at the last time point and not the more distant history.
- There is evidence that rating histories show momentum and stickiness which violates the Markov assumption (Lando and Skodeberg (2002)).

Properties of Markov chains

- The Markov chain is stationary if, for all $t \geq 1$ and rating states j, k,

$$\mathbb{P}(R_t = k \mid R_{t-1} = j) = \mathbb{P}(R_1 = k \mid R_0 = j).$$

- In this case we can define the transition matrix $\boldsymbol{P} = (p_{jk})$ with elements $p_{jk} = \mathbb{P}(R_t = k \mid R_{t-1} = j)$, for any $t \geq 1$.
- The Chapman-Kolmogorov equations say that

$$\mathbb{P}(R_t = k \mid R_{t-2} = j) = \sum_{l \in S} p_{jl} p_{lk} \,.$$

- An implication of this is that the matrix of transition probabilities over two time steps is given by $\boldsymbol{P}^2 = \boldsymbol{P} \times \boldsymbol{P}$.
- It is not clear how a matrix of transition probabilities for a fraction of a time period can be computed (one would need continuous-time chains).
- Estimators of transition probabilities from rating history are available, see Section 10.2.2 of the book.

10.3 Structural models of default

10.3.1 The Merton model

- Merton's model (1974) is the prototype of all firm value models.
- Consider firm with stochastic asset-value (V_t), financing itself by equity (i.e. by issuing shares) and debt.
- Assume that debt consists of single zero coupon bond with face or nominal value B and maturity T.
- Denote by S_t and B_t the value at time $t \leq T$ of equity and debt so that

$$V_t = S_t + B_t, \quad 0 \leq t \leq T.$$

- Assume that default occurs if the firm misses a payment to its debt holders and hence only at T.

Equity and debt as contingent claims on assets

- At T we have two possible cases:

  1) $V_T > B$. In that case the debtholders receive B; shareholders receive residual value $S_T = V_T - B$, and there is no default.
  2) $V_T \leq B$. In that case the firm cannot meet its financial obligations, and shareholders hand over control to the bondholders, who liquidate the firm; hence we have $B_T = V_T$, $S_T = 0$.

- In summary we obtain

$$S_T = (V_T - B)^+$$
$$B_T = \min(V_T, B) = B - (B - V_T)^+ .$$

- The value of equity at T equals the pay-off of a European call option on V_T with exercise price equal to B.
- The value of the debt at T equals the nominal value of debt minus the pay-off of a European put option on V_T.

- The option interpretation explains certain conflicts of interest between shareholders and bondholders.
- For example, shareholders have more interest in the firm taking on risky projects/investments since the value of an option increases with the volatility of the underlying security.
- Bondholders have a short position on the firm's assets and would like to see the volatility reduced.

The asset value process

It is assumed that asset value (V_t) follows a diffusion of the form

$$dV_t = \mu_V V_t dt + \sigma_V V_t dW_t$$

for constants $\mu_V \in \mathrm{R}$, $\sigma_V > 0$, and a Brownian motion $(W_t)_{t \geq 0}$, so that

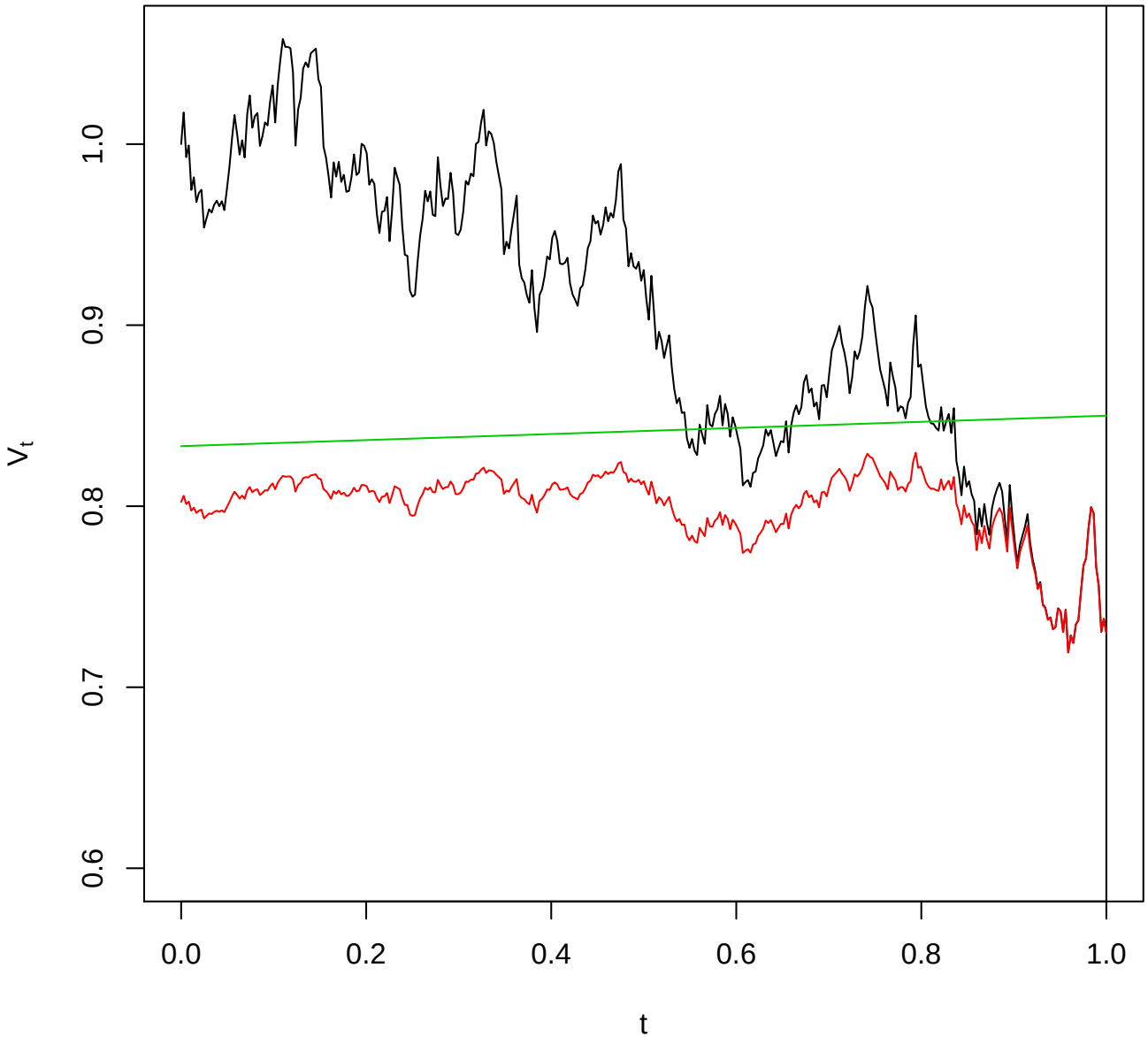
$$V_T = V_0 \exp\Big((\mu_V - \frac{1}{2}\sigma_V^2)T + \sigma_V W_T \Big);$$

in particular $\ln V_T \sim \mathrm{N}(\ln V_0 + (\mu_V - \frac{1}{2}\sigma_V^2)T,\ \sigma_V^2 T)$. The default probability is thus

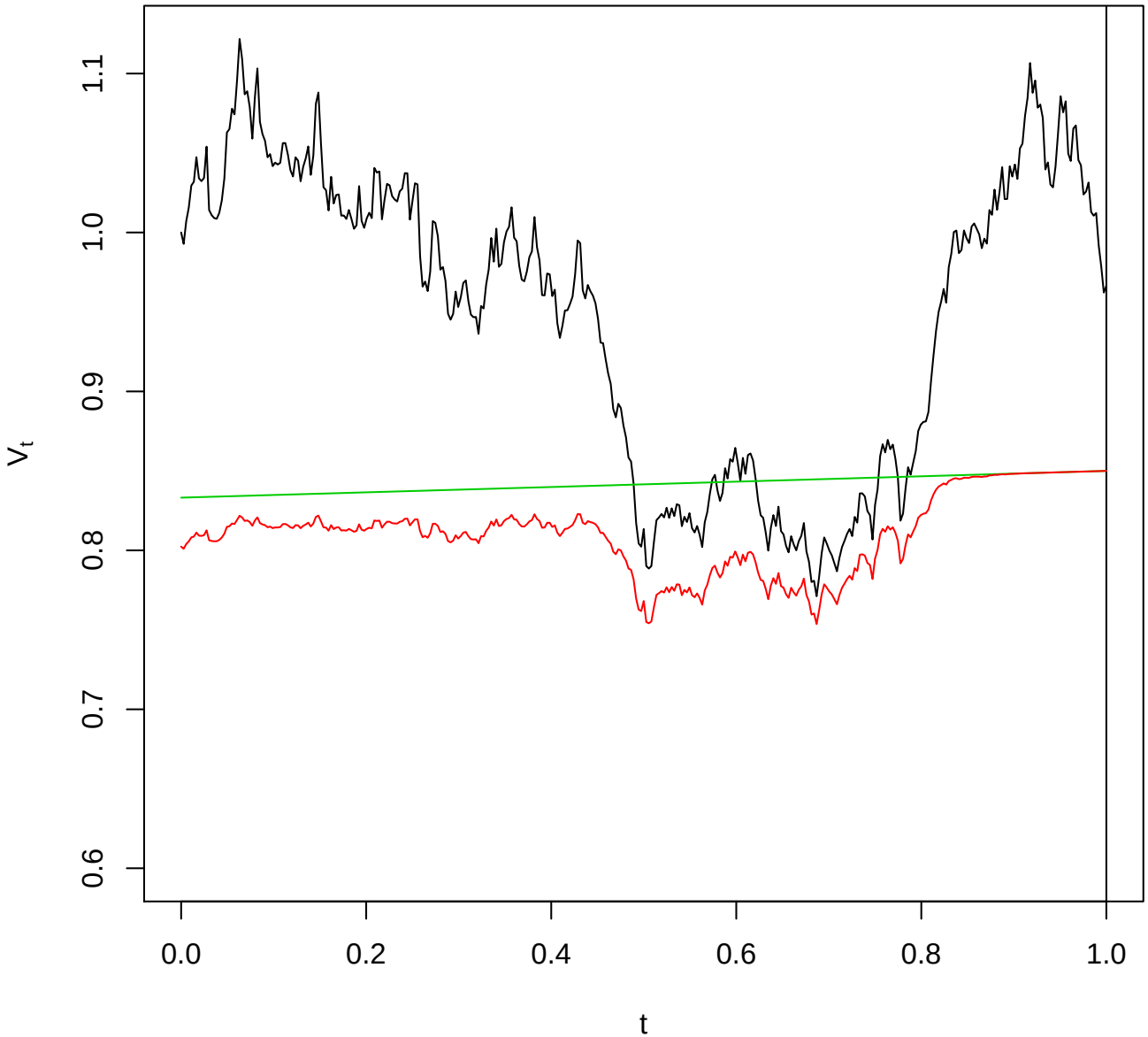
$$\mathbb{P}(V_T \leq B) = \mathbb{P}(\ln V_T \leq \ln B) = \Phi\left(\frac{\ln \frac{B}{V_0} - (\mu_V - \frac{1}{2}\sigma_V^2)T}{\sigma_V \sqrt{T}} \right); \quad (52)$$

it is increasing in B and σ_V (for $V_0 > B$) and decreasing in V_0 and μ_V.

A default path



A non-default path



10.3.2 Pricing in Merton's model

- Under some technical assumptions we can price equity and debt using the Black–Scholes formula.
- The assumptions are that:

  1) The risk-free rate is deterministic and equal to $r \geq 0$.

  2) The asset-value process (V_t) is independent of the debt level B.

  3) The asset value (V_t) can be traded on a frictionless market.

- Recall that equity is a call option on the asset value (V_t). Hence Black–Scholes formula yields

$$S_t = C^{\mathsf{BS}}(t, V_t; \sigma_V, r, T, B) := V_t \Phi(d_{t,1}) - B e^{-r(T-t)} \Phi(d_{t,2}),$$

  where the arguments are given by

$$d_{t,1} = \frac{\ln \frac{V_t}{B} + (r + \frac{1}{2}\sigma_V^2)(T-t)}{\sigma_V \sqrt{T-t}}, \quad d_{t,2} = d_{t,1} - \sigma_V \sqrt{T-t}.$$

Pricing of debt

- The price at $t \leq T$ of a default-free zero-coupon bond with maturity T and a face value of one equals

$$p_0(t,T) = \exp(-r(T-t)).$$

- The value of the firm's debt equals the difference between the value of default-free debt and a put option on (V_t) with strike B, i.e.

$$B_t = Bp_0(t,T) - P^{\mathsf{BS}}(t, V_t; r, \sigma_V, B, T).$$

- The Black–Scholes formula for European puts now yields

$$B_t = p_0(t,T)B\Phi(d_{t,2}) + V_t\Phi(-d_{t,1}). \tag{53}$$

- The path of (B_t) is shown on the previous plots. The value of default-free debt $Bp_0(t,T)$ is shown as a green curve.

Risk-neutral and physical default probabilities

- Under the risk-neutral measure $\mathbb{Q}$ the process (V_t) satisfies the SDE $\mathrm{d}V_t = rV_t\,\mathrm{d}t + \sigma_V V_t\,\mathrm{d}\tilde{W}_t$ for a standard $\mathbb{Q}$-Brownian motion $\tilde{W}$.
- The drift μ_V is replaced by the risk-free interest rate r.
- Hence the risk-neutral default probability is given by

$$q = \mathbb{Q}(V_T \leq B) = \Phi\left(\frac{\ln B - \ln V_0 - (r - \frac{1}{2}\sigma_V^2)T}{\sigma_V\sqrt{T}}\right).$$

- Comparison with physical default probability $p = \mathbb{P}(V_T \leq B)$ yields

$$q = \Phi\left(\Phi^{-1}(p) + \frac{\mu_V - r}{\sigma_V}\sqrt{T}\right). \tag{54}$$

- The correction term $(\mu_V - r)/\sigma_V$ equals the Sharpe ratio of the firm's assets (a popular measure of the risk premium earned by the firm).
- The formula is sometimes applied in practice to go from physical to risk-neutral default probabilities.

Credit spreads in Merton's model

- The credit spread measures the difference between the (continuously compounded) yield of a default-free zero coupon bond $p_0(t,T)$ and a defaultable zero coupon bond $p_1(t,T)$, i.e.

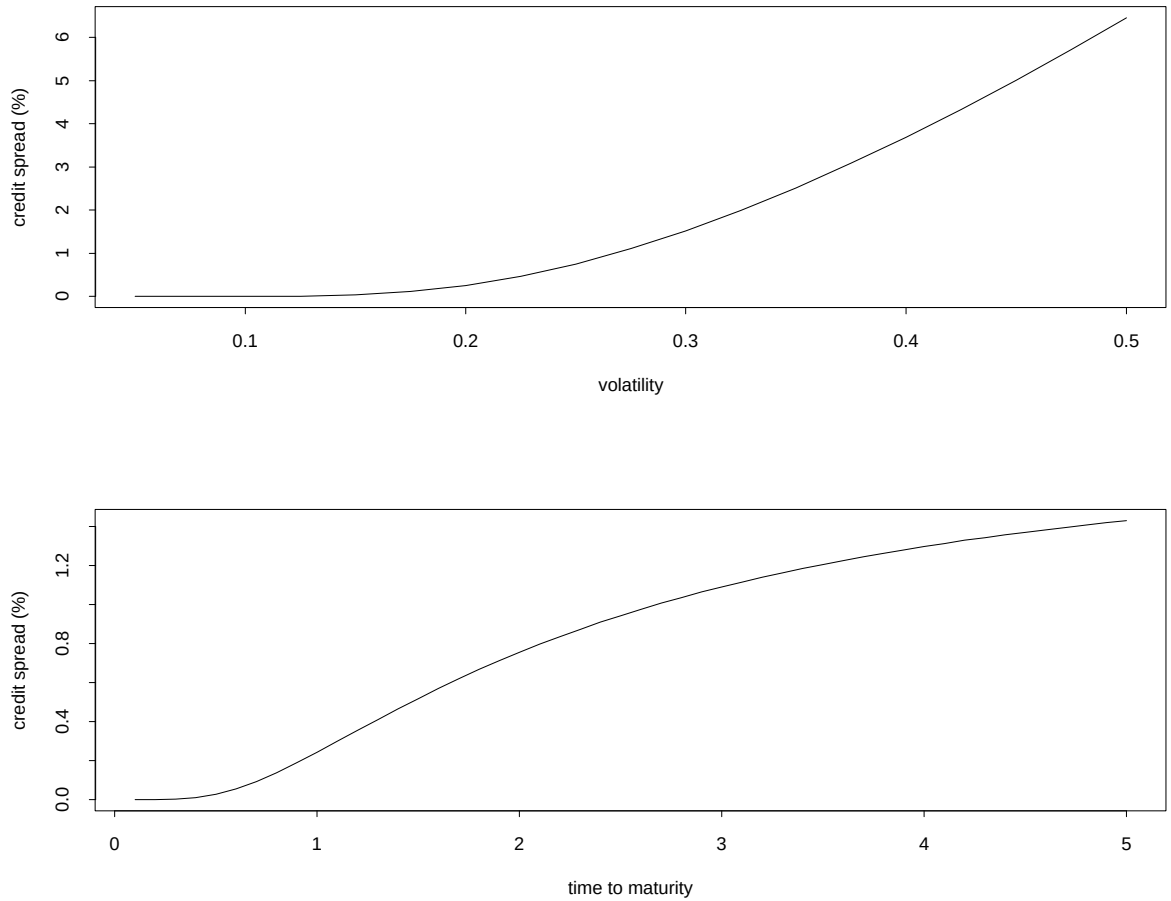
$$\begin{aligned} c(t,T) &= \frac{-1}{T-t}\left(\ln p_1(t,T) - \ln p_0(T-t)\right) \\ &= \frac{-1}{T-t}\ln\frac{p_1(t,T)}{p_0(t,T)}\,. \end{aligned}$$

- In Merton's model we have $p_1(t,T) = \frac{1}{B}B_t$ and hence

$$c(t,T) = \frac{-1}{(T-t)}\ln\left(\Phi(d_{t,2}) + \frac{V_t}{Bp_0(t,T)}\,\Phi(-d_{t,1})\right). \tag{55}$$

- For a fixed time to maturity $c(t,T)$ depends only on σ_V and on the ratio $Bp_0(t,T)/V_t$ (a measure of indebtedness of the firm).
- In line with economic intuition it is increasing in both quantities.

Illustration of credit spreads in Merton's model



Credit spread $c(t,T)$ (%) as function of σ_V (top) and time to maturity $T-t$ (bottom) for fixed debt to firm value ratio 0.6. In upper picture $T-t=2$; in lower picture $\sigma_V = 0.25$.

10.4 Bond and CDS pricing in hazard rate models

10.4.1 Hazard rate models

- These are the simplest reduced-form credit risk models.
- A hazard rate model is a model in which the distribution of the default time of an obligor is directly specified by a hazard function without modelling the mechanism by which default occurs.
- To set up a hazard rate model we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random default time τ defined on this space, i.e. an $\mathcal{F}$-measurable rv taking values in $[0, \infty]$.
- We denote the df of τ by $F(t) = \mathbb{P}(\tau \leq t)$ and the tail or survival function by $\bar{F}(t) = 1 - F(t)$; we assume that $\mathbb{P}(\tau = 0) = F(0) = 0$, and that $\bar{F}(t) > 0$ for all $t < \infty$.

- The jump or default indicator process (Y_t) associated with τ is

$$Y_t = I_{\{\tau \leq t\}}, \quad t \geq 0. \tag{56}$$

- (Y_t) is a right-continuous process which jumps from 0 to 1 at the default time τ.
- $1 - Y_t = I_{\{\tau > t\}}$ is the survival indicator of the firm at time t.

Definition 10.1 (cumulative hazard and hazard function)
The function $\Gamma(t) = -\ln(\bar{F}(t))$ is called the cumulative hazard function of the random time τ. If F is absolutely continuous with density f, the function

$$\gamma(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{\bar{F}(t)} = -\frac{\mathrm{d}}{\mathrm{d}t} \ln(\bar{F}(t))$$

is called the hazard function of τ.

- The hazard function $\gamma(t)$ gives the hazard rate at t, which is a measure of the instantaneous risk of default at t, given survival up to time t.

- We can represent the survival function of τ by

$$\bar{F}(t) = \exp\Big(- \int_0^t \gamma(s)\,\mathrm{d}s \Big). \tag{57}$$

- We may show that

$$\lim_{h\to 0} \frac{1}{h}\mathbb{P}(\tau \le t+h \mid \tau > t) = \frac{1}{\bar{F}(t)} \lim_{h\to 0} \frac{F(t+h)-F(t)}{h} = \gamma(t).$$

Example 10.2 (Weibull distribution)
For illustrative purposes we determine the hazard function for the Weibull distribution with df $F(t) = 1 - \exp(-\lambda t^\alpha)$ for parameters $\lambda, \alpha > 0$. Differentiation yields

$$f(t) = \lambda\alpha t^{\alpha-1}\exp(-\lambda t^\alpha) \text{ and } \gamma(t) = \lambda\alpha t^{\alpha-1}.$$

In particular, γ is decreasing in t if $\alpha < 1$ and increasing if $\alpha > 1$. For $\alpha = 1$ (exponential distribution) the hazard rate equals the constant λ.

Introducing filtrations

- Filtrations model information available to investors over time.
- A filtration $(\mathcal{F}_t)$ on $(\Omega, \mathcal{F})$ is an increasing family $\{\mathcal{F}_t : t \geq 0\}$ of sub-σ-algebras of $\mathcal{F}$: $\mathcal{F}_t \subset \mathcal{F}_s \subset \mathcal{F}$ for $0 \leq t \leq s < \infty$.
- $\mathcal{F}_t$ represents the state of knowledge of an observer at time t. $A \in \mathcal{F}_t$ means that at time t we can determine if A has occurred.
- In this section we assume that only observable quantity is the default indicator (Y_t) associated with τ. The appropriate filtration is $(\mathcal{H}_t)$ with

$$\mathcal{H}_t = \sigma(\{Y_u : u \leq t\}), \tag{58}$$

  the default history up to and including time t.
- τ is a $(\mathcal{H}_t)$-stopping time, since $\{\tau \leq t\} = \{Y_t = 1\} \in \mathcal{H}_t$ for all $t \geq 0$.
- In order to study bond and CDS pricing in hazard rate models we need to compute conditional expectations with respect to the σ-algebra $\mathcal{H}_t$.

A useful result

Lemma 10.3

Let τ be a default time with jump indicator process $Y_t = I_{\{\tau \leq t\}}$ and natural filtration $(\mathcal{H}_t)$. Then, for any integrable rv X and any $t \geq 0$, we have

$$\mathbb{E}(I_{\{\tau>t\}} X \mid \mathcal{H}_t) = I_{\{\tau>t\}} \frac{\mathbb{E}(I_{\{\tau>t\}} X)}{\mathbb{P}(\tau > t)}. \tag{59}$$

This result can be used to determine conditional survival probabilities. For $t < T$, applying (59) with $X := I_{\{\tau>T\}}$ we get

$$\mathbb{P}(\tau > T \mid \mathcal{H}_t) = I_{\{\tau>t\}} \exp\Big(- \int_t^T \gamma(s)\, \mathrm{d}s \Big), \quad t < T\,. \tag{60}$$

Martingale property of jump indicator process

Proposition 10.4

The process (M_t) defined as

$$M_t = Y_t - \int_0^t I_{\{\tau > u\}} \gamma(u)\,\mathrm{d}u, \quad t \geq 0$$

is an $(\mathcal{H}_t)$-martingale, that is $\mathbb{E}(M_s \mid \mathcal{H}_t) = M_t$ for all $0 \leq t \leq s < \infty$.

10.4.2 Risk-neutral pricing revisited

- According to the first fundamental theorem of asset pricing, a model for security prices is arbitrage free if and (essentially) only if it admits at least one equivalent martingale measure $\mathbb{Q}$.
- When building a model for pricing derivatives it is a natural shortcut to model the objects of interest—such as interest rates and default times—directly, under a martingale measure $\mathbb{Q}$.

Physical ($\mathbb{P}$) vs risk-neutral ($\mathbb{Q}$) measure: An example

- Consider a defaultable bond with principal 1 and maturity $T = 1$y. In case of a default (real world probability $p = 0.01$), the recovery rate is $R = 60\%$. The risk-free interest rate is $r = 0.05$. Moreover, assume the bond's current price to be $V_0 = 0.941$ $(t = 0)$.
- The expected discounted value of the bond is

$$\frac{1}{1+r}(1 \cdot (1-p) + R \cdot p) = \frac{1}{1.05}(0.99 + 0.6p) = 0.949$$

  which is $> V_0$ since investors demand a premium for bearing the bond's default risk.
- Here, $\mathbb{Q}$ is determined by specifying a q such that

$$\frac{1}{1+r}(1 \cdot (1-q) + R \cdot q) = V_0.$$

  This implies $q = 0.03$ which is greater than $p = 0.01$; the larger value reflects the risk premium.

Martingale modelling

- So-called martingale modelling is particularly convenient if the value H of the underlying assets at some maturity date T is exogenously given, as in the case of zero-coupon bonds.
- The underlying asset at time $t < T$ can be computed as the conditional expectation under $\mathbb{Q}$ of the discounted value at maturity via the risk-neutral pricing rule

$$V_t = \mathbb{E}^{\mathbb{Q}}\left(e^{-\int_t^T r_s \mathrm{d}s} H \mid \mathcal{F}_t\right). \tag{61}$$

- Model parameters are determined using the requirement that at time $t = 0$ the model price should coincide with the market price of the security; this is known as calibration to market data.

Pros and cons of Martingale modelling

- Martingale modelling ensures that the resulting model is arbitrage free, which is important for pricing many different securities simultaneously.
- The approach is frequently adopted in default-free term structure models and in reduced-form models for credit-risky securities.
- Martingale modelling has two drawbacks.

  1) Historical information is largely useless in estimating model parameters.

  2) Realistic models for pricing credit derivatives are typically incomplete.

- An arbitrage-free market is complete if and only if there is exactly one equivalent martingale measure.
- In incomplete markets there may be more than one equivalent martingale measure. It will generally not be possible to find a replicating strategy for a derivative (one cannot eliminate all risk by dynamic hedging).

10.4.3 Bond pricing

- It suffices to consider zero-coupon bonds.
- We use martingale modelling and work directly under some martingale measure $\mathbb{Q}$.
- We assume that under $\mathbb{Q}$ the default time τ is a random time with deterministic risk-neutral hazard function $\gamma^{\mathbb{Q}}(t)$.
- The information available to investors at time t is given by the sigma algebra $\mathcal{H}_t = \sigma(\{Y_u : u \leq t\})$.
- We take interest rates and recovery rates to be deterministic.
- The percentage loss given default is denoted by $\delta \in (0, 1)$.
- The continuously compounded interest rate is denoted by $r(t) \geq 0$.
- The price of the default-free zero-coupon bond with maturity $T \geq t$ is $p_0(t, T) = \exp(-\int_t^T r(s)\,\mathrm{d}s)$.

Analysing the payments

- The payments of a defaultable zero-coupon bond can be represented as a combination of a survival claim that pays one unit at the maturity date T and a recovery payment in case of default.
- The survival claim has pay-off $I_{\{\tau>T\}}$.
- Recall from (60) that

$$\mathbb{Q}(\tau > T \mid \mathcal{H}_t) = I_{\{\tau>t\}} \exp\Big(- \int_t^T \gamma^{\mathbb{Q}}(s)\,\mathrm{d}s\Big)$$

  and define $R(t) = r(t) + \gamma^{\mathbb{Q}}(t)$.
- Then the price of a survival claim at time t equals

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}(p_0(t,T) I_{\{\tau>T\}} \mid \mathcal{H}_t) &= \exp\Big(- \int_t^T r(s)\,\mathrm{d}s\Big)\mathbb{Q}(\tau > T \mid \mathcal{H}_t) \\ &= I_{\{\tau>t\}} \exp\Big(- \int_t^T R(s)\,\mathrm{d}s\Big)\,. \qquad (62)\end{aligned}$$

- Note that for $\tau > t$, this can be viewed as the price of a default-free zero-coupon bond with adjusted interest rate $R(t) > r(t)$.
- A similar relationship between defaultable and default-free bond prices can be established in many reduced-form credit risk models.

Recovery models

1) Recovery of Treasury (RT).
   - The RT model was proposed by Jarrow and Turnbull (1995).
   - If default occurs at some point in time $\tau \leq T$, the owner of the defaulted bond receives $(1 - \delta_\tau)$ units of the default-free zero-coupon bond $p_0(\cdot, T)$ at time τ, where $\delta_\tau \in [0, 1]$ models the percentage loss given default.
   - At maturity T the holder of the defaultable bond therefore receives the payment $I_{\{\tau > T\}} + (1 - \delta_\tau) I_{\{\tau \leq T\}}$.

2) Recovery of Face Value (RF).

- Under RF, if default occurs at $\tau \leq T$, the holder of the bond receives a recovery payment of size $(1 - \delta_\tau)$ immediately at the default time τ.
- Note that even with deterministic loss given default and deterministic interest rates, the value at maturity of the recovery payment is random as it depends on the exact timing of default.

RF is slightly more realistic; RT is slightly easier to analyse.

Pricing recovery payment under RT

- The value of the recovery payment at the maturity date T is

$$(1-\delta)I_{\{\tau \leq T\}} = (1-\delta) - (1-\delta)I_{\{\tau > T\}}.$$

- Using (62), the value of the recovery payment at time $t < T$ is hence

$$(1-\delta)p_0(t,T) - (1-\delta)I_{\{\tau > t\}} \exp\Big(-\int_t^T R(s)\,\mathrm{d}s\Big).$$

- Hence the value of the bond is

$$p_1(t,T) = (1-\delta)p_0(t,T) + \delta I_{\{\tau>t\}} \exp\Big(-\int_t^T R(s)\,\mathrm{d}s\Big).$$

Pricing recovery payment under RF

- Under the RF-hypothesis the recovery payment takes the form $(1-\delta)I_{\{\tau\leq T\}}$ where the payment occurs directly at time τ.
- A payments of this form is a payment-at-default claim.
- The value of the recovery payment at time $t \leq T$ equals

$$\mathbb{E}^{\mathbb{Q}}\Big((1-\delta)I_{\{t<\tau\leq T\}} \exp\Big(-\int_t^\tau r(s)\,\mathrm{d}s\Big)\,\Big|\,\mathcal{H}_t\Big).$$

- Using (59) we may show that

$$\begin{aligned}&\mathbb{E}^{\mathbb{Q}}\Big((1-\delta)I_{\{t<\tau\leq T\}} \exp\Big(-\int_t^\tau r(s)\,\mathrm{d}s\Big)\,\Big|\,\mathcal{H}_t\Big)\\ &= (1-\delta)I_{\{\tau>t\}} \int_t^T \gamma^{\mathbb{Q}}(s) \exp\Big(-\int_t^s R(u)\,\mathrm{d}u\Big)\,\mathrm{d}s.\end{aligned}$$

10.4.4 CDS pricing

First we recall payment flows. We write $\tau = \tau_C$ and consider the following contract:

- Premium payments.
  - These are due at times $0 < t_1 < \cdots < t_N$ measured in years.
  - If $\tau > t_k$, A pays a premium of size $x^*(t_k - t_{k-1})$ at t_k, where x^* denotes the fair swap spread.
  - After τ premium payments stop, no initial payment.
- Default payment.
  - If $\tau < t_N = T$, B makes a default payment δ at τ.
  - Sometimes B receives an accrued premium payment of size $x^*(\tau - t_k)$ for $\tau \in (t_k, t_{k-1})$. We ignore this feature for simplicity.

Valuing the premium leg

- The premium leg consists of a set of survival claims.
- Introduce a function of x given by

$$\begin{aligned} & V_t^{\text{prem}}(x;\gamma^{\mathbb{Q}}) \\ &= \mathbb{E}^{\mathbb{Q}}\bigg(\sum_{k:\, t_k>t} \exp\bigg(-\int_t^{t_k} r(u)\,\mathrm{d}u \bigg) x(t_k - t_{k-1}) I_{\{\tau>t_k\}} \mid \mathcal{H}_t \bigg) \\ &= x \sum_{k:\, t_k>t} p_0(t,t_k)(t_k - t_{k-1}) \mathbb{Q}(\tau > t_k \,|\, \mathcal{H}_t), \end{aligned}$$

  which is easily computed using $\mathbb{Q}(\tau > t_k \,|\, \mathcal{H}_t) = \exp(-\int_t^{t_k} \gamma^{\mathbb{Q}}(s)\,\mathrm{d}s)$.
- We obtain

$$V_t^{\text{prem}}(x;\gamma^{\mathbb{Q}}) = I_{\{\tau>t\}}\, x \sum_{k:\, t_k>t} (t_k - t_{k-1}) \exp\bigg(-\int_t^{t_k} R(u)\,\mathrm{d}u \bigg).$$

Valuing the default leg

- The default payment leg is a typical payment-at-default claim.
- We obtain

$$\begin{aligned} & V_t^{\mathsf{def}}(\gamma^{\mathbb{Q}}) \\ &= \mathbb{E}^{\mathbb{Q}}\Big(\delta I_{\{t<\tau\leq t_N\}} \exp\Big(-\int_t^{\tau} r(s)\,\mathrm{d}s\Big)\,\Big|\,\mathcal{H}_t\Big) \\ &= I_{\{\tau>t\}}\,\delta \int_t^{t_N} \gamma^{\mathbb{Q}}(s) \exp\Big(-\int_t^{s} R(u)\,\mathrm{d}u\Big)\,\mathrm{d}s. \end{aligned}$$

The fair CDS spread

- The fair CDS spread x_t^* quoted for the contract at time t is chosen such that the value of the contract is equal to zero.

- The equation $V_t^{\text{prem}}(x_t^*;\gamma^{\mathbb{Q}}) = V_t^{\text{def}}(\gamma^{\mathbb{Q}})$ yields

$$x_t^* = I_{\{\tau>t\}} \frac{\delta \int_t^{t_N} \gamma^{\mathbb{Q}}(s) \exp\Big(- \int_t^s R(u)\,\mathrm{d}u\Big)\,\mathrm{d}s}{\sum_{k:\, t_k>t}(t_k - t_{k-1}) \exp\Big(- \int_t^{t_k} R(s)\,\mathrm{d}s\Big)}\,. \tag{63}$$

Model calibration

- We have to calibrate our model to the available market information. Hence we have to determine the implied risk-neutral hazard function $\gamma^{\mathbb{Q}}(t)$, which ensures that the fair CDS spreads implied by the model equal the spreads quoted in the market.
- Suppose that the market information at time $t = 0$ consists of the fair spread x^* of one CDS with maturity t_N.
- In that case $\gamma^{\mathbb{Q}}(s)$ is taken constant: for all $s \geq 0$, $\gamma^{\mathbb{Q}}(s) = \bar{\gamma}^{\mathbb{Q}}$ for some $\bar{\gamma}^{\mathbb{Q}} > 0$.

- $\bar{\gamma}^{\mathbb{Q}}$ has to solve the equation

$$x^* \sum_{k=1}^{N} p_0(0,t_k)(t_k - t_{k-1}) e^{-\bar{\gamma}^{\mathbb{Q}} t_k} = \delta \bar{\gamma}^{\mathbb{Q}} \int_0^{t_N} p_0(0,t) e^{-\bar{\gamma}^{\mathbb{Q}} t} \, \mathrm{d}t.$$

- There is a unique solution.
- If we observe spreads for several CDSs on the same reference entity but with different maturities, a constant function is not sufficient. Instead one typically uses piecewise constant or linear hazard functions.
- A first approximation to the implied hazard rate is given by $\bar{\gamma}^{\mathbb{Q}} \approx x^*/\delta$.
- This approximation implies that the one-year default probability satisfies $\mathbb{Q}(\tau \leq 1) = 1 - e^{-\bar{\gamma}^{\mathbb{Q}}} \approx \bar{\gamma}^{\mathbb{Q}} \approx x^*/\delta$.

11 Portfolio credit risk management

11.1 Threshold models

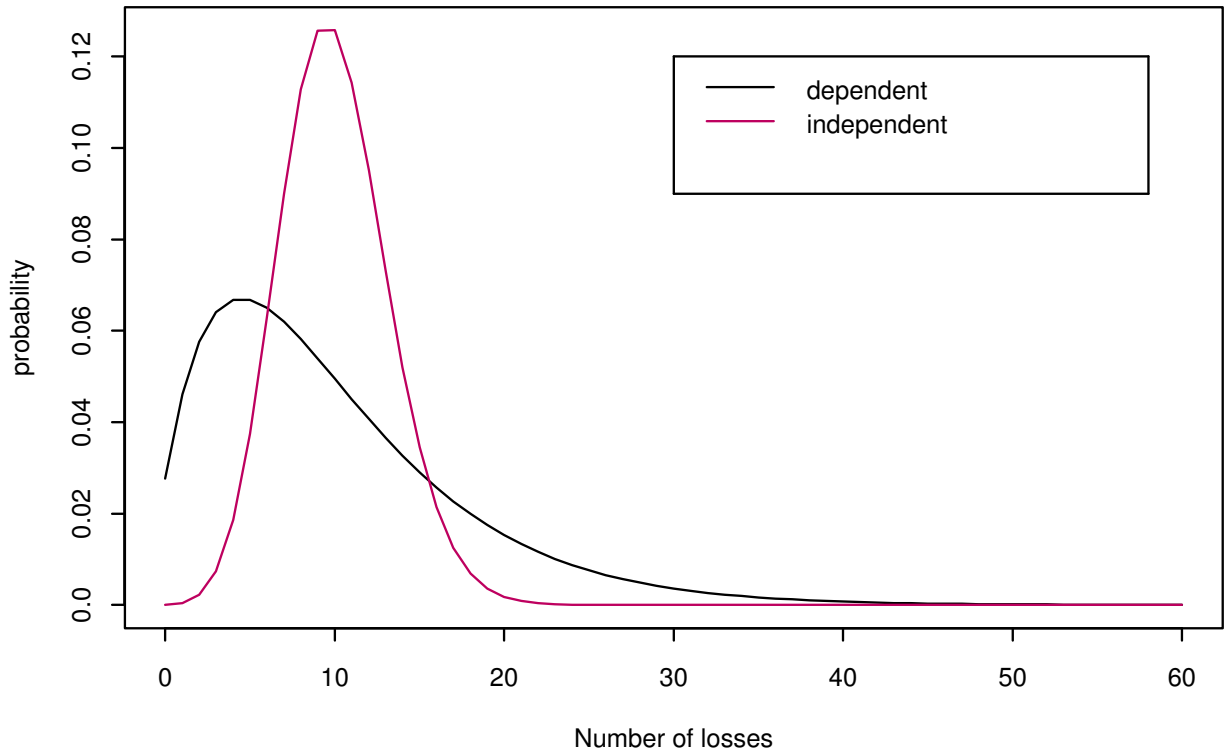
11.2 Mixture models

11.3 Statistical inference for portfolio credit models

Importance of default dependence

Dependence between defaults (and downgrades) is a key issue in credit risk management. There are two main sources of dependence between defaults:

- Dependence caused by common factors (for example, interest rates and changes in economic growth) affecting all obligors
- Default of company A may have direct impact on default probability of company B and vice versa because of direct business relations, a phenomenon known as contagion



Comparison of the loss distribution of a homogeneous portfolio of 1000 loans with a default probability of $p_1 = \cdots = p_{1000} = 1\%$ assuming (i) independent defaults and (ii) a default correlation of $\rho(Y_i, Y_j) = 0.5\%$. Case (ii) can be considered as roughly representative for BB-rated loans.

11.1 Threshold models

11.1.1 Notation for one-period portfolio models

- Consider portfolio of m firms and time horizon $T = 1$ (say one year).
- For $1 \leq i \leq m$, let R_i be a state indicator for obligor i at time T taking values in the set $\{0, 1, \ldots, n\}$; we interpret the value 0 as default and non-zero values as states of increasing credit quality. At time $t = 0$ obligors are assumed to be in some non-default state.
- Mostly we will concentrate on the binary outcomes of default and non-default. We write Y_i for the default indicator variables so that $Y_i = 1 \iff R_i = 0$ and $Y_i = 0 \iff R_i > 0$.
- The random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_m)'$ is a vector of default indicators for the portfolio and $p(\boldsymbol{y}) = \mathbb{P}(Y_1 = y_1, \ldots, Y_m = y_m)$, $\boldsymbol{y} \in \{0, 1\}^m$, is its joint probability function; the marginal default probabilities are denoted by $p_i = \mathbb{P}(Y_i = 1)$, $i = 1, \ldots, m$.

- Default or event correlation. Noting that

$$\mathrm{var}(Y_i) = \mathbb{E}(Y_i^2) - p_i^2 = \mathbb{E}(Y_i) - p_i^2 = p_i - p_i^2,$$

we obtain, for firms i and j with $i \neq j$, the formula

$$\rho(Y_i, Y_j) = \frac{\mathbb{E}(Y_i Y_j) - p_i p_j}{\sqrt{(p_i - p_i^2)(p_j - p_j^2)}}. \tag{64}$$

- Let the rv $M := \sum_{i=1}^m Y_i$ denote the number of defaulted obligors at T.
- The actual loss if company i defaults is modelled by the random quantity $\delta_i e_i$, where e_i represents the overall exposure to company i and $0 \leq \delta_i \leq 1$ represents the LGD.
- We denote the overall portfolio loss by $L := \sum_{i=1}^m \delta_i e_i Y_i$.
- It is possible to set up different credit risk models leading to the same multivariate distribution for $\boldsymbol{R}$ or $\boldsymbol{Y}$. We call two models with state vectors $\boldsymbol{R}$ and $\tilde{\boldsymbol{R}}$ (or $\boldsymbol{Y}$ and $\tilde{\boldsymbol{Y}}$) equivalent if $\boldsymbol{R} \stackrel{\mathrm{d}}{=} \tilde{\boldsymbol{R}}$ (or $\boldsymbol{Y} \stackrel{\mathrm{d}}{=} \tilde{\boldsymbol{Y}}$).

11.1.2 Threshold models and copulas

Definition 11.1
Let $\boldsymbol{X} = (X_1, \ldots, X_m)'$ be an m-dimensional random vector and let $D \in \mathbb{R}^{m \times n}$ be a deterministic matrix with elements d_{ij} such that, for every i, the elements of the ith row form a set of increasing thresholds satisfying $d_{i1} < \cdots < d_{in}$. Augment these thresholds by setting $d_{i0} = -\infty$ and $d_{i(n+1)} = \infty$ for all obligors and then set

$$R_i = j \iff d_{ij} < X_i \leq d_{i(j+1)}, \quad j \in \{0, \ldots, n\},\ i \in \{1, \ldots, m\}.$$

Then $(\boldsymbol{X}, D)$ is said to define a threshold model for $\boldsymbol{R} = (R_1, \ldots, R_m)'$.

- $\boldsymbol{X}$ are the critical variables and the ith row of D contains the critical thresholds for firm i.
- Default occurs if $X_i \leq d_{i1}$ so that the default probability of company i is given by $p_i = F_{X_i}(d_{i1})$.

- When working with a default-only model we simply write $d_i = d_{i1}$ and denote the threshold model by $(\boldsymbol{X}, \boldsymbol{d})$.
- Default correlation and asset correlation. It is important to distinguish the default correlation $\rho(Y_i, Y_j)$ of two firms $i \neq j$ from the correlation of the critical variables X_i and X_j.
- Since the critical variables are often interpreted in terms of asset values, the latter correlation is often referred to as asset correlation.
- For given default probabilities, $\rho(Y_i, Y_j)$ is determined by $\mathbb{E}(Y_i Y_j)$ according to (64), and in a threshold model $\mathbb{E}(Y_i Y_j) = \mathbb{P}(X_i \leq d_{i1},\ X_j \leq d_{j1})$, so default correlation depends on the joint df of X_i and X_j.
- If $\boldsymbol{X}$ is multivariate normal, as in many models used in practice, the correlation of X_i and X_j determines the copula of their joint distribution and hence the default correlation.
- If two threshold models lead to the same state/default probabilities and if the critical variables have the same copula, they are equivalent.

11.1.3 Gaussian threshold models

Multivariate Merton model:

- Assume that the multivariate asset-value process $\boldsymbol{V}_t = (V_{t,1}, \ldots, V_{t,m})'$ follows an m-dimensional GBM with drift vector $\boldsymbol{\mu}_V = (\mu_1, \ldots, \mu_m)'$, vector of volatilities $\boldsymbol{\sigma}_V = (\sigma_1, \ldots, \sigma_m)'$ and correlation matrix P.
- This means that $(\boldsymbol{V}_t)$ solves the stochastic differential equations

$$\mathrm{d}V_{t,i} = \mu_i V_{t,i}\,\mathrm{d}t + \sigma_i V_{t,i}\,\mathrm{d}W_{t,i}, \quad i = 1, \ldots, m,$$

  for correlated BMs with correlation $\rho(W_{t,i}, W_{t,j}) = \rho_{ij}$, $t \geq 0$.
- For all i the asset value $V_{T,i}$ is of the form

$$V_{T,i} = V_{0,i} \exp\big((\mu_i - \tfrac{1}{2}\sigma_i^2)T + \sigma_i W_{T,i}\big),$$

  where $\boldsymbol{W}_T \sim \mathrm{N}_m(\boldsymbol{0}, TP)$.
- In its basic form the Merton model is a default-only model where the firm defaults if $V_{T,i} \leq B_i$ and B_i is the liability of firm i.

- Writing $\boldsymbol{B} = (B_1, \ldots, B_m)'$ the threshold model representation is $(\boldsymbol{V}_T, \boldsymbol{B})$.
- The multivariate Merton model is equivalent to the model $(\boldsymbol{X}, \boldsymbol{d})$ with

$$X_i := \frac{\ln V_{T,i} - \ln V_{0,i} - (\mu_i - \frac{1}{2}\sigma_i^2)T}{\sigma_i \sqrt{T}},$$
$$d_i := \frac{\ln B_i - \ln V_{0,i} - (\mu_i - \frac{1}{2}\sigma_i^2)T}{\sigma_i \sqrt{T}}.$$

- The transformed variables satisfy $\boldsymbol{X} \sim \mathrm{N}_m(\boldsymbol{0}, P)$ and their copula is the Gauss copula C_P^{Ga}.

Gaussian threshold models in practice

- In practice it is usual to start directly with threshold models of the form $(\boldsymbol{X}, \boldsymbol{d})$ with $\boldsymbol{X} \sim \mathrm{N}_m(\boldsymbol{0}, P)$.

- There are two practical challenges:

  1) calibration of $\boldsymbol{d}$ (or, in the case of a multi-state model, the threshold matrix D) in line with exogenously given default and transition probabilities;

  2) calibration of the correlation matrix P in a parsimonious way.

- The problem of embedding state transition probabilities in a threshold matrix D has already been discussed. In a default-only model we set $d_i = \Phi^{-1}(p_i)$ for $i = 1, \ldots, m$.

Factor models

- In its most general form P has $m(m-1)/2$ distinct parameters.
- m is typically large and it is important to use a more parsimonious parametrization of this matrix based on a factor model.

- Factor models also lend themselves to economic interpretation and the factors are commonly interpreted as country and industry effects.
- We assume that

$$X_i = \sqrt{\beta_i}\tilde{F}_i + \sqrt{1-\beta_i}\varepsilon_i, \tag{65}$$

  where $\tilde{F}_i$ and $\varepsilon_1, \ldots, \varepsilon_m$ are independent standard normal variables, and where $0 \leq \beta_i \leq 1$ for all i.
- In this formulation $\tilde{F}_i$ are the systematic variables, which are correlated, and ε_i are idiosyncratic variables.
- It follows that β_i can be viewed as a measure of the systematic risk of X_i: that is, the part of the variance of X_i which is explained by the systematic variable.
- The systematic variables are assumed to be of the form $\tilde{F}_i = \boldsymbol{a}_i'\boldsymbol{F}$ where $\boldsymbol{F}$ is a vector of common factors satisfying $\boldsymbol{F} \sim \mathrm{N}_p(\boldsymbol{0}, \Omega)$ with $p < m$, and where Ω is a correlation matrix.

- These factors typically represent country and industry effects.
- The assumption that $\mathrm{var}(\tilde{F}_i) = 1$ means that $\boldsymbol{a}_i' \Omega \boldsymbol{a}_i = 1$ for all i.
- Since $\mathrm{var}(X_i) = 1$ and since $\tilde{F}_i$ and $\varepsilon_1, \ldots, \varepsilon_m$ are independent and standard normal, the asset correlations in this model are given by

$$\rho(X_i, X_j) = \mathrm{cov}(X_i, X_j) = \sqrt{\beta_i \beta_j}\, \mathrm{cov}(\tilde{F}_i, \tilde{F}_j) = \sqrt{\beta_i \beta_j}\, \boldsymbol{a}_i' \Omega \boldsymbol{a}_j \,.$$

- In order to set up the model we have to determine $\boldsymbol{a}_i$ and β_i for each obligor and Ω, with the additional constraint that $\boldsymbol{a}_i' \Omega \boldsymbol{a}_i = 1$ for all i.
- Since Ω has $p(p-1)/2$ parameters, the loading vectors $\boldsymbol{a}_i$ and coefficients β_i have collectively $mp + m$ parameters, and we are applying m constraints, this gives $mp + p(p-1)/2$ parameters.

11.2 Mixture models

11.2.1 Bernoulli mixture models

- In a mixture model the default risk of an obligor is assumed to depend on a set of common factors, usually interpreted as macroeconomic variables, which are also modelled stochastically.
- Given a realization of the factors, defaults of individual firms are assumed to be independent.
- Dependence between defaults stems from the dependence of individual default probabilities on the set of common factors.
- Bernoulli mixture models provide a way of capturing the dependence between Bernoulli events (i.e. defaults/non-defaults).
- They can be extended to multinomial mixture models to capture dependent migrations in a rating system.

Definition 11.2 (Bernoulli mixture model)
Given some $p < m$ and a p-dimensional random vector $\mathbf{\Psi} = (\Psi_1, \dots, \Psi_p)'$, the default indicator vector $\mathbf{Y}$ follows a Bernoulli mixture model with factor vector $\mathbf{\Psi}$ if there are functions $p_i : \mathbb{R}^p \to (0,1)$, such that conditional on $\mathbf{\Psi}$ the components of $\mathbf{Y}$ are independent Bernoulli rvs with $\mathbb{P}(Y_i = 1 \mid \mathbf{\Psi} = \boldsymbol{\psi}) = p_i(\boldsymbol{\psi})$.

The conditional independence given factors makes these models relatively easy to analyse. For $\mathbf{y} = (y_1, \dots, y_m)'$ in $\{0,1\}^m$ we get

$$\mathbb{P}(\mathbf{Y} = \mathbf{y} \mid \mathbf{\Psi} = \boldsymbol{\psi}) = \prod_{i=1}^{m} p_i(\boldsymbol{\psi})^{y_i} (1 - p_i(\boldsymbol{\psi}))^{1-y_i}$$

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}) = \int_{\mathbb{R}^p} \prod_{i=1}^{m} p_i(\boldsymbol{\psi})^{y_i} (1 - p_i(\boldsymbol{\psi}))^{1-y_i} g(\boldsymbol{\psi}) \, \mathrm{d}\boldsymbol{\psi},$$

where $g(\boldsymbol{\psi})$ is the probability density of the factors. The default probabilities are given by $p_i = \mathbb{E}(Y_i = 1) = \mathbb{E}(p_i(\mathbf{\Psi}))$.

- Consider the portfolio loss $L = \sum_{i=1}^m e_i \delta_i Y_i$ in the case where the exposures e_i and LGDs δ_i are deterministic.
- It is difficult to compute the df F_L of L.
- However, it is easy to use the conditional independence of the defaults to show that the *Laplace–Stieltjes transform* of F_L is for $t \in \mathbb{R}$ given by

$$\begin{aligned}\hat{F}_L(t) = \mathbb{E}(e^{-tL}) &= \mathbb{E}\Big(\mathbb{E}(e^{-t\sum_{i=1}^m e_i\delta_i Y_i} \mid \boldsymbol{\Psi})\Big) \\ &= \mathbb{E}\bigg(\prod_{i=1}^m \mathbb{E}(e^{-te_i\delta_i Y_i} \mid \boldsymbol{\Psi})\bigg) \\ &= \mathbb{E}\bigg(\prod_{i=1}^m (p_i(\boldsymbol{\Psi})e^{-te_i\delta_i} + 1 - p_i(\boldsymbol{\Psi}))\bigg)\end{aligned}$$

  which can be obtained by integrating over distribution of factors $\boldsymbol{\Psi}$.
- This is useful for: sampling losses from model with importance sampling; approximating probability mass function using Fourier inversion.

11.2.2 Threshold models as mixture models

- Although the mixture models of this section seem, at first glance, to be different in structure to the threshold models, it is important to realize that the majority of useful threshold models, including all the examples we have given, can be represented as Bernoulli mixture models.
- In a threshold model default occurs for counterparty i if a critical variable X_i lies below a critical threshold d_i.
- Moreover X_i follows a linear factor model

$$X_i = \sqrt{\beta_i}\boldsymbol{a}_i'\boldsymbol{F} + \sqrt{1-\beta_i}\varepsilon_i, \quad \text{where}$$

  - $\boldsymbol{F} \sim \mathrm{N}_p(\mathbf{0}, \Omega)$ is a random vector of normally distributed common economic factors;
  - $0 \leq \beta_i \leq 1$ and $\mathrm{var}(\boldsymbol{a}_i'\mathbf{F}) = 1$;
  - $\varepsilon_1, \ldots, \varepsilon_m$ are iid standard normal and are also independent of $\boldsymbol{F}$.

- We will write the Gaussian threshold model as a Bernoulli mixture model with factor vector $\boldsymbol{\Psi} = -\boldsymbol{F}$. (This makes the conditional default probabilities increasing in the factors for positive $\boldsymbol{a}_i$.)
- Conditioning on $\boldsymbol{\Psi} = -\boldsymbol{F}$, the vector $\boldsymbol{X}$ is multivariate normally distributed with a diagonal covariance matrix and therefore the components of $\boldsymbol{X}$ are conditionally independent.
- The conditional default probabilities are

$$\begin{aligned} p_i(\boldsymbol{\psi}) = \mathbb{P}(Y_i = 1 \,|\, \boldsymbol{\Psi} = \boldsymbol{\psi}) &= \mathbb{P}(X_i \leq d_i \,|\, \boldsymbol{\Psi} = \boldsymbol{\psi}) \\ &= \mathbb{P}(X_i \leq d_i \,|\, \boldsymbol{F} = -\boldsymbol{\psi}) \\ &= \mathbb{P}(\sqrt{1-\beta_i}\varepsilon_i \leq d_i + \sqrt{\beta_i}\mathbf{a}_i'\boldsymbol{\psi}) \\ &= \Phi\left(\frac{d_i + \sqrt{\beta_i}\mathbf{a}_i'\boldsymbol{\psi}}{\sqrt{1-\beta_i}}\right) \\ &= \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\mathbf{a}_i'\boldsymbol{\psi}}{\sqrt{1-\beta_i}}\right). \end{aligned}$$

11.2.3 Poisson mixture models and CreditRisk+

- Since default is typically a rare event, it is possible to approximate Bernoulli indicator rvs for default with Poisson rvs and Bernoulli mixture models with Poisson mixture models.
- By choosing independent gamma distributions for the economic factors $\mathbf{\Psi}$, we obtain a tractable model known as CreditRisk+, proposed by Credit Suisse Financial Products in 1997.
- Assume that, given the factors $\mathbf{\Psi}$, the default indicators $Y_1, \ldots, Y_m$ for a particular time horizon are conditionally independent Bernoulli variables satisfying $\mathbb{P}(Y_i = 1 \mid \mathbf{\Psi} = \boldsymbol{\psi}) = p_i(\boldsymbol{\psi})$.
- Moreover assume that the distribution of $\mathbf{\Psi}$ is such that the conditional default probabilities $p_i(\boldsymbol{\psi})$ tend to be very small.
- The Y_i variables can be approximated by conditionally independent Poisson variables $\tilde{Y}_i$ satisfying $\tilde{Y}_i \mid \mathbf{\Psi} = \boldsymbol{\psi} \sim \text{Poi}(p_i(\boldsymbol{\psi}))$.

- This follows because

$$\mathbb{P}(\tilde{Y}_i = 0 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = e^{-p_i(\boldsymbol{\psi})} \approx 1 - p_i(\boldsymbol{\psi}),$$
$$\mathbb{P}(\tilde{Y}_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = p_i(\boldsymbol{\psi}) e^{-p_i(\boldsymbol{\psi})} \approx p_i(\boldsymbol{\psi}).$$

- The portfolio loss $L = \sum_{i=1}^m e_i \delta_i Y_i$ can be approximated by $\tilde{L} = \sum_{i=1}^m e_i \delta_i \tilde{Y}_i$.
- It is possible for a company to "default more than once" in the approximating Poisson model, albeit with a very low probability.
- In CreditRisk+ the parameter $\lambda_i(\boldsymbol{\Psi})$ of the conditional Poisson distribution for firm i is assumed to take the form

$$\lambda_i(\boldsymbol{\Psi}) = k_i \boldsymbol{w}_i' \boldsymbol{\Psi} \tag{66}$$

for $k_i > 0$, non-negative weights $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{ip})'$ satisfying $\sum_j w_{ij} = 1$, and p independent $\mathrm{Ga}(\alpha_j, \beta_j)$-distributed factors $\Psi_1, \ldots, \Psi_p$.

- The parameters are set to be $\alpha_j = \beta_j = \sigma_j^{-2}$ for $\sigma_j > 0$ and $j = 1, \ldots, p$.

- This parametrization of the gamma variables ensures that we have $\mathbb{E}(\Psi_j) = 1$ and $\mathrm{var}(\Psi_j) = \sigma_j^2$.
- It is easy to verify that

$$\mathbb{E}(\tilde{Y}_i) = \mathbb{E}(\mathbb{E}(\tilde{Y}_i \mid \boldsymbol{\Psi})) = \mathbb{E}(\lambda_i(\boldsymbol{\Psi})) = k_i \mathbb{E}(\boldsymbol{w}_i' \boldsymbol{\Psi}) = k_i,$$

  so k_i is the expected number of defaults for obligor i in the time period.
- The assumptions in CreditRisk+ make it possible to compute the distribution of the number of defaults and the aggregate portfolio loss fairly explicitly using techniques for compound distributions and mixture distributions that are well known in actuarial mathematics.

Distribution of the number of defaults

In CreditRisk+ we have that given $\boldsymbol{\Psi} = \boldsymbol{\psi}$, $\tilde{Y}_i \sim \mathrm{Poi}(k_i \boldsymbol{w}_i' \boldsymbol{\psi})$, which implies that the distribution of the number of defaults $\tilde{M} := \sum_{i=1}^m \tilde{Y}_i$ satisfies

$$\tilde{M} \mid \boldsymbol{\Psi} = \boldsymbol{\psi} \sim \mathrm{Poi}\left(\sum_{i=1}^m k_i \boldsymbol{w}_i' \boldsymbol{\psi} \right). \tag{67}$$

- This uses the fact that the sum of independent Poisson variables is also Poisson with a rate parameter given by the sum of the rate parameters
- To compute the unconditional distribution of $\tilde{M}$ we require a well-known result on mixed Poisson distributions.

Proposition 11.3
If the rv N is conditionally Poisson with a gamma-distributed rate parameter $\Lambda \sim \mathrm{Ga}(\alpha, \beta)$, then N has a negative binomial distribution, $N \sim \mathrm{NB}(\alpha, \beta/(\beta+1))$.

In the case when $p = 1$ we may apply this result directly to (67) to deduce that $\tilde{M}$ has a negative binomial distribution. The general result is:

Proposition 11.4
$\tilde{M}$ is distributed as a sum of p independent negative binomial rvs.

Distribution of the aggregate loss

- To obtain a tractable model, exposures are discretized in CreditRisk+ using the concept of exposure bands.
- The LGD is subsumed in the exposure by multiplying the actual exposure by a typical value for the LGD for an obligor with the same credit rating.
- The losses arising from the individual obligors are of the form $\tilde{L}_i = e_i \tilde{Y}_i$ where the e_i are known (LGD-adjusted) exposures.
- For all i, the exposure e_i is discretized in units of an amount ϵ so that e_i is replaced by a value $\ell_i \epsilon \geq e_i$ where ℓ_i is a positive integer multiplier.
- Exposure bands $b = 1, \ldots, n$ are defined corresponding to the distinct values $\ell^{(1)}, \ldots, \ell^{(n)}$ for the multipliers so that obligors are grouped in exposure bands according to the values of their discretized exposures.
- It is then possible to derive the distribution of the aggregate loss $\tilde{L} = \sum_{i=1}^{m} \ell_i \epsilon \tilde{Y}_i$.

Theorem 11.5

Let $\tilde{L}$ represent the aggregate loss in the general p-factor CreditRisk+ model with exposures discretized into exposure bands as described above. Then the following hold.

i) The Laplace–Stieltjes transform of the df of $\tilde{L}$ is given by

$$\hat{F}_{\tilde{L}}(s) = \prod_{j=1}^{p} \left(1 + \sigma_j^2 \sum_{i=1}^{m} k_i w_{ij} \left(1 - \sum_{b=1}^{n} e^{-s\epsilon\ell^{(b)}} q_{jb}\right)\right)^{-\sigma_j^{-2}}, \quad (68)$$

where $q_{jb} = \sum_{i \in s_b} k_i w_{ij} / \sum_{i=1}^{m} k_i w_{ij}$ for $b = 1, \ldots, n$.

ii) The distribution of $\tilde{L}$ has the structure $\tilde{L} \stackrel{d}{=} \sum_{j=1}^{p} Z_j$ where the Z_j are independent variables that follow a compound negative binomial distribution. More precisely, it holds that $Z_j \sim \mathrm{CNB}(\sigma_j^{-2}, \theta_j, G_{X_j})$ with $\theta_j = (1 + \sigma_j^2 \sum_{i=1}^{m} k_i w_{ij})^{-1}$ and G_{X_j} the df of a multinomial random variable X_j taking the value $\epsilon\ell^{(b)}$ with probability q_{jb}.

11.3 Statistical inference for portfolio credit models

11.3.1 Industry factor models

- Recall that portfolio models in industry often take the form of a Gaussian threshold model $(\boldsymbol{X}, \boldsymbol{d})$ with $\boldsymbol{X} \sim \mathrm{N}_m(\mathbf{0}, P)$, where the random vector $\boldsymbol{X}$ contains the critical variables, the deterministic vector $\boldsymbol{d}$ contains the critical default thresholds and P is the so-called asset correlation matrix, which is estimated with the help of a factor model for $\boldsymbol{X}$.
- Industry models generally separate the calibation of the vector $\boldsymbol{d}$ (or the threshold matrix D in a multi-state model) and the calibration of the factor model for $\boldsymbol{X}$.
- In a default-only model the threshold d_i is usually set at $d_i = \Phi^{-1}(p_i)$ where p_i is an estimate of the default probability for obligor i for the time period in question (generally one year).

- The default probability may be estimated in different ways: for larger corporates it may be estimated using credit ratings or using a firm-value approach, such as the Moody's public-firm EDF model; for retail obligors it may be estimated on the basis of credit scores.
- Recall that the factor model for $\boldsymbol{X}$ takes the form

$$X_i = \sqrt{\beta_i}\tilde{F}_i + \sqrt{1-\beta_i}\varepsilon_i, \quad i = 1, \ldots, m, \tag{69}$$

  where $\tilde{F}_i$ and $\varepsilon_1, \ldots, \varepsilon_m$ are independent standard normal variables, and where $0 \leq \beta_i \leq 1$ for all i.
- The systematic variables $\tilde{F}_i$ are assumed to be of the form $\tilde{F}_i = \boldsymbol{a}_i'\boldsymbol{F}$ where $\boldsymbol{F}$ is a vector of common factors satisfying $\boldsymbol{F} \sim \mathrm{N}_p(\boldsymbol{0}, \Omega)$ with $p < m$, and where Ω is a correlation matrix.
- The factors typically represent country and industry effects.
- The assumption that $\mathrm{var}(\tilde{F}_i) = 1$ implies that $\boldsymbol{a}_i'\Omega\boldsymbol{a}_i = 1$ for all i.

- Different industry models use different data for $\boldsymbol{X}$ to calibrate the factor model (69).
- The Moody's Analytics Global Correlation or GCorr model has sub-models for many different kinds of obligor including public corporate firms, private firms, small and medium enterprises (SMEs), retail customers and sovereigns. Huang et al. (2012)
- The sub-model for public firms (GCorr Corporate) is calibrated using data on weekly asset value returns, where asset values are determined as part of the public-firm EDF methodology.
- In the CreditMetrics framework weekly equity returns are viewed as a proxy for asset returns and used to estimate the factor model.
- We sketch a generic procedure for estimating a factor model for corporates where the factors have country and industry-sector interpretations.

Estimating a credit risk factor model

- We assume that we have a high-dimensional multivariate time series $(\boldsymbol{X}_t)_{1\leq t\leq n}$ of asset returns (or other proxy data for changing credit quality) over a period of time in which stationarity can be assumed.
- We also assume that each component time series has been scaled to have mean zero and variance one.

1) We first fix the structure of the factor vector $\boldsymbol{F}$ so that, for example, the first block of components might represent country factors and the second block of components might represent industry factors. We then assign vectors of factor weights $\boldsymbol{a}_i$ to each obligor based on our knowledge of the companies. The elements of $\boldsymbol{a}_i$ may simply consist of ones and zeros if the company can be clearly identified with a single country and industry, but may also consist of weights if the company has significant activity in more than one country or sector.

2) We then use cross-sectional estimation techniques to estimate the factor values $\boldsymbol{F}_t$ at each time point t. Effectively the factor estimates $\hat{\boldsymbol{F}}_t$ are constructed as weighted sums of the $X_{t,i}$ data for obligors i that are exposed to each factor. One way of achieving this is to construct a matrix A with rows $\boldsymbol{a}_i$ and then to estimate a fundamental factor model of the form $\boldsymbol{X}_t = A\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t$ at each time point t.

- We have a regression model

$$\boldsymbol{X}_t = A\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \tag{70}$$

  where $\boldsymbol{X}_t \in \mathbb{R}^m$ are the return data, $A \in \mathbb{R}^{m\times p}$ is a known matrix of factor loadings, $\boldsymbol{F}_t \in \mathbb{R}^p$ are the factors to be estimated and $\boldsymbol{\varepsilon}_t$ are errors with diagonal covariance matrix Υ.

- Note that the components of the error vector $\boldsymbol{\varepsilon}_t$ can not generally be assumed to have equal variance, so that (70) is a regression problem with so-called heteroskedastic errors.

- Unbiased estimators of the factors $\boldsymbol{F}_t$ may be obtained by forming the ordinary least squares (OLS) estimates

$$\hat{\boldsymbol{F}}_t^{\text{OLS}} = (A'A)^{-1}A'\boldsymbol{X}_t\,.$$

- Since the errors are heteroskedastic, slightly more efficient estimators can be obtained by using the method of generalized least squares (GLS).

3) The raw factor estimates form a multivariate time series of dimension p. We standardize each component series to have mean zero and variance one to obtain $(\hat{\boldsymbol{F}}_t)_{1\leq t\leq n}$ and calculate the sample covariance matrix of the standardized factor estimates, which serves as our estimate of Ω.

4) We then scale the vectors of factor weights $\boldsymbol{a}_i$ so that the conditions $\boldsymbol{a}_i'\hat{\Omega}\boldsymbol{a}_i = 1$ are met for each obligor.

5) Time series of estimated systematic variables for each obligor are then constructed by calculating $\hat{\tilde{F}}_{t,i} = \boldsymbol{a}_i'\hat{\boldsymbol{F}}_t$ for $t = 1, \ldots, n$.

6) Finally we estimate the β_i parameters by performing a time series regression of $X_{t,i}$ on $\hat{\tilde{F}}_{t,i}$ for each obligor.

Note that the accurate estimation of the β_i in the last step is particularly important (as it effects tail behaviour). The estimate of β_i is the so-called R-squared of the time series regression model in Step 6 and will be largest for the firms whose credit-quality changes are best explained by systematic factors.

17 Introduction to counterparty risk

17.1 Introduction

17.2 Credit value adjustments

17.1 Introduction

- A substantial part of all derivative transactions is carried out over the counter and there is no central clearing counterparty to guarantee fulfilment of the contractual obligations.
- These trades are subject to the risk that a contracting party defaults during the transaction, thus affecting the cash flows that are actually received by the other party. This is known as counterparty credit risk.
- Counterparty risk received a lot of attention during the financial crisis of 2007-2009 as some of the institutions heavily involved in derivative transactions experienced worsening credit quality or—in the case of Lehman Brothers—even a default event.
- Counterparty risk management is now a key issue for all financial institutions and the focus of many new regulatory developments.

Example of Interest-Rate Swap

- Two parties A and B agree to exchange a series of interest payments on a given nominal amount of money for a given period.
- A receives payments at a fixed interest rate and makes floating payments at a rate equal to the three-month LIBOR rate.
- Suppose that A defaults at time τ_A before the maturity of the contract.
- If interest rates have risen relative to their value at inception of contract:
  - The fixed interest payments have decreased in value and the value of the contract has increased for B.
  - The default of A constitutes a loss for B; the loss size depends on the term structure of interest rates at τ_A.
- If interest rates have fallen relative to their value at $t = 0$:
  - The fixed payments have increased in value so that the swap has a negative value for B.

- B will still has to pay the value of the contract into the bankruptcy pool, and there is no upside for B in A's default.

- If B defaults first the situation is reversed: falling rates lead to a counterparty-risk-related loss for A.

Management of counterparty risk

- Counterparty risk has to be taken into account in pricing and valuation. This has led to the notion of credit value adjustments (CVA).
- Counterparty risk needs to be controlled using risk-mitigation techniques such as netting and collateralization.
- Under a netting agreement the value of all derivatives transactions between A and B is computed and only the aggregated value is subject to counterparty risk; since offsetting transactions cancel each other out, this has the potential to reduce counterparty risk substantially.
- Under a collateralization agreement the parties exchange collateral (cash and securities) that serves as a pledge for the receiver. The value of the collateral is adjusted dynamically to reflect changes in the value of the underlying transactions.

17.2 Credit value adjustments

General definition. The price (for the protection buyer) satisfies

$$\begin{aligned} \text{True price} = \; & \text{(counterparty) risk-free price} \\ & - \text{ adjustment for default of seller (CVA)} \\ & + \text{ adjustment for default of buyer (DVA)} , \end{aligned}$$

where CVA and DVA stand for Credit Value Adjustment and Debt Value Adjustment respectively.

General adjustment formulas

Denote by V_t the market value of the CDS (assuming that B and S are default-free), by $\tau = \min\{\tau_R, \tau_S, \tau_B\}$ the first default time and by $\xi \in \{R, S, B\}$ the identity of first defaulting firm. Recall that $x^+ = \max(x, 0)$ and $x^- = -\min(x, 0)$ and denote by $D(0, t)$ the discount factor over the period $[0, t]$ (with constant interest rate, $D(0, t) = e^{-rt}$).

It can be shown that

$$\text{CVA} = \mathbb{E}^{\mathbb{Q}}(I_{\{\tau<T\}} I_{\{\xi=S\}} D(0, \tau)\delta^S V_\tau^+)$$
$$\text{DVA} = \mathbb{E}^{\mathbb{Q}}(I_{\{\tau<T\}} I_{\{\xi=B\}} D(0, \tau)\delta^B V_\tau^-)$$

Comments.

- CVA gives loss of B due to premature default of S; DVA gives loss of S due to premature default of B.
- The value adjustments involve an option on the market value $V = (V_t)_{t \leq T}$ of the swap with strike $K = 0$ (a call for the CVA and a put for the DVA).
- Similar formula holds if V is the market value of another derivative such as an interest swap or even a reinsurance contract.
- DVA is a bit problematic: a worsening credit quality of B leads to an accounting profit for B.

A simplified formula

In order to evaluate the CVA and DVA formulas one needs a model with stochastic credit spreads that takes dependence between the default of S, B and the market value V of the CDS into account (a dynamic portfolio credit risk model). Markets often work with a simpler formula that assumes that the default of S and B and V are independent:

$$\mathrm{CVA}^{\text{indep}} = \delta^S \int_0^T \bar{F}_B(t) D(0,t) \mathbb{E}^{\mathbb{Q}}(V_t^+) f_S(t)\,\mathrm{d}t,$$

$$\mathrm{DVA}^{\text{indep}} = \delta^B \int_0^T \bar{F}_S(t) D(0,t) \mathbb{E}^{\mathbb{Q}}(V_t^-) f_B(t)\,\mathrm{d}t.$$

Here f_S is the density of τ_S and $\bar{F}_B$ resp $\bar{F}_S$ is the survival function of τ_B resp τ_S.

Comments.

- In order to evaluate the simplified formula one only needs to determine the marginal distribution of τ_S and τ_B and the so-called expected exposures $\mathbb{E}^{\mathbb{Q}}(V_t^+)$ and $\mathbb{E}^{\mathbb{Q}}(V_t^-)$.
- The independence assumption underlying the simplified value adjustment formula between the price of the CDS on R, that is V_t, and the default event of S and B is often unrealistic; in practice this is known as wrong way risk.

  Examples:

  - CDS on a financial institution: given that S defaults it is quite likely that credit quality of R is low.
  - Reinsurance.

For further reading on counterparty risk see Gregory (2012).

References

Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999), Coherent measures of risk, *Mathematical Finance*, 9, 203–228.

Balkema, A. A. and de Haan, L. (1974), Residual life time at great age, *The Annals of Probability*, 2, 792–804.

CME SPAN: Standard Portfolio Analysis of Risk (2010), `www.cmegroup.com/` Chicago Mercantile Exchange.

D'Agostino, R. B. and Stephens, M. A. (1986), Goodness-of-fit techniques, Dekker.

Demarta, S. and McNeil, A. J. (2005), The t Copula and Related Copulas, *International Statistical Review*, 73(1), 111–129.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), Modelling Extremal Events for Insurance and Finance, Berlin: Springer.

Embrechts, Paul and Hofert, Marius (2013), Statistical inference for copulas in high dimensions: A simulation study, *ASTIN Bulletin*, 43(2), 81–95, doi:10.1017/asb.2013.6.

Föllmer, H. and Schied, A. (2002), Convex measures of risk and trading constraints, *Finance and Stochastics*, 6, 429–447.

Genest, C., Ghoudi, K., and Rivest, L. (1995), A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, 82, 543–552.

Genest, C. and Rivest, L. (1993), Statistical inference procedures for bivariate Archimedean copulas, *Journal of the American Statistical Association*, 88, 1034–1043.

Gnedenko, B. V. (1943), Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics*, 44, 423–453.

Gneiting, T. (2011), Making and Evaluating Point Forecasts, *Journal of the American Statistical Association*, 106(494), 746–762.

Gregory, J. (2012), Counterparty Credit Risk and Credit Value Adjustment: A Continuing Challenge for Global Financial Markets, Chichester: John Wiley and Sons ltd.

Higham, Nick (2002), Computing the nearest correlation matrix – A problem from finance, *IMA Journal of Numerical Analysis*, 22, 329–343.

Hofert, Marius and Scherer, Matthias (2011), CDO pricing with nested Archimedean copulas, *Quantitative Finance*, 11(5), 775–787, doi:http://dx.doi.org/10.1080/14697680903508479.

Huang, Jing-Zhi, Lanfranconi, M., Patel, N., and Pospisil, L. (2012-12), Modeling credit correlations: an overview of the Moody's Analytics GCorr Model, tech. rep., Moody's Analytics.

Iman, R. L. and Conover, W. J. (1982), A distribution-free approach to inducing rank correlation among input variables, *Communications in Statistics – Simulation and Computation*, 11(3), 311–334.

Jarrow, R. A. and Turnbull, S. M. (1995), Pricing Derivatives on Financial Securities Subject to Credit Risk, *The Journal of Finance*, 50(1), 53–85.

Joe, H. and Xu, J. J. (1996), The Estimation Method of Inference Functions for Margins for Multivariate Models, *Technical Report 166, Department of Statistics, University of British Columbia*.

Joenssen, D. W. and Vogel, J. (2014), A power study of goodness-of-fit tests for multivariate normality implemented in R, *Journal of Statistical Computation and Simulation*, 84, 1055–1078.

Kim, G., Silvapulle, M. J., and Silvapulle, P. (2007), Comparison of semiparametric and parametric methods for estimating copulas, *Computational Statistics & Data Analysis*, 51, 2836–2850.

Kloman, H. F. (1990), Risk management agonists, *Risk Analysis*, 10, 201–205.

Kolmogorov, A. N. (1933), Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin: Ergebnisse der Mathematik.

Lando, D. and Skodeberg, T. (2002), Analyzing Rating Transitions and Rating Drift with Continuous Observations, *Journal of Banking and Finance*, 26, 423–444.

Li, D. X. (2000), On Default Correlation: A Copula Function Approach, *The Journal of Fixed Income*, 9(4), 43–54.

Mashal, R. and Zeevi, A. (2002), Beyond correlation: Extreme co-movements between financial assets, Preprint, Columbia Business School.

McNeil, A. J. and Frey, R. (2000), Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach, *Journal of Empirical Finance*, 7, 271–300.

McNeil, A. J., Frey, R., and Embrechts, P. (2005), Quantitative Risk Management: Concepts, Techniques and Tools, Princeton: Princeton University Press.

McNeil, A. J., Frey, R., and Embrechts, P. (2015), Quantitative Risk Management: Concepts, Techniques and Tools, 2nd, Princeton: Princeton University Press.

Ou, S. (2013), Annual Default Study: Corporate Default and recovery Rates, 1920-2012, Technical document, Moody's Investor Service.

Pickands, J. (1975), Statistical inference using extreme order statistics, *The Annals of Statistics*, 3, 119–131.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992), Numerical Recipes in C, Cambridge: Cambridge University Press.

RiskMetrics (1996), RiskMetrics Technical Document, 3rd, J.P. Morgan, New York.

Schönbucher, P. J. and Schubert, D. (2001), Copula-Dependent Default Risk in Intensity Models, `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=301968` (2009-12-30).

Shreve, S. E. (2008), Don't blame the quants, Available at `www.forbes.com/2008shreve.html`.

Sibuya, M. (1960), Bivariate extreme statistics, *Annals of the Institute of Statistical Mathematics*, 11, 195–210.

Smith, R. L. (1987), Estimating Tails of Probability Distributions, *The Annals of Statistics*, 15, 1174–1207.