1 Introduction

Model Uncertainty

Model uncertainty refers to the uncertainty about the accuracy of a model. It results from imprecise and idealized assumptions, which, to some degree, have to be made in every modeling framework.

Different Types of Risk

Market risk; Credit risk; Liquidity risk [including (1)Market liquidity risk, (2)Funding liquidity risk]; Operational risk; Underwriting risk; Model risk (an example of Knightian uncertainty)

What does managing risks involve?

- Determine enough buffer capital to absorb losses (for regulatory purposes and economic capital purposes).
- Making sure portfolios are well diversified.
- Optimizing portfolios according to risk-return considerations.

Recent developments and concerns

High frequency trading; Algorithmic trading; Commodities trading; Systemic risk

Basel Accord

The first Basel Accord (Basel I), 1988

- Only addressed credit risk.
- Fairly coarse measurement of risk. Loans were divided into 3 categories only, counterparties being governments, regulated banks and others.
- Risk weighting identical for all corporate borrowers, independent of their credit rating.
- Unsatisfactory treatment of derivatives.
- Banks pushed to be allowed to use netting
- Amendment to Basel 1 in 1996: (1) Standardized model for market risk and internal VaR-based models for more sophisticated banks. (2) Coarseness problem for credit risk remained.

The second Basel Accord (Basel II), 2004

Three pillar concept:

Pillar 1 Minimal capital charge: Requirements for the calculation of the regulatory capital to ensure that a bank holds sufficient capital for its credit risk in the banking book, its market risk in the trading book and operational risk (which was added as a new class of risk)

Pillar 2 Supervisory review process: Local regulators review the checks and balances put in place for capital adequacy assessments, ensure that banks have adequate regulatory capital and perform stress tests of a bank's capital adequacy

Pillar 3 Market discipline: Banks are required to make risk management processes more transparent

Basel 2.5, 2009

- CDOs provided opportunities for regulatory arbitrage (transferring credit risk from the capital-intensive banking book to the less-capitalized trading book)
- The aim of Basel 2.5 was to address the build up of risk in the trading book
- It included (1) stressed VaR calculations of positions in the trading book; (2) incremental risk charge due to possible defaults and rating changes; (3) Exposure to securitizations in the trading book was subjected to new capital charges

The third Basel Accord (Basel III), 2010

Intends to increase bank liquidity and decrease bank leverage.
Five extensions:

(1) It increases the quality and amount of capital by changing the definition of key capital ratios and allowing countercyclical adjustments to these ratios in crises

(2) It strengthens the framework for counterparty credit risk in derivatives trading with incentives to use central counterparties (exchanges)

(3) It introduces a leverage ratio to prevent excessive leverage (a way to multiply gains/losses by buying more of an asset with borrowed capital)

(4) It introduces various ratios that ensure that banks have sufficient funding liquidity

(5) It forces systemically important financial institutions (SIFIs) to hold even more risk capital

Basel IV (anticipated)

- Would require more stringent capital requirements
- Emphasizes simpler or standardized models in place of bank internal models
- Requires more detailed disclosure of reserves and other financial statistics

From Solvency I to II

- Solvency I: Rather coarse rules-based framework calling for companies to have a minimum guarantee fund. Simple robust system, easy to understand, inexpensive to monitor. However, it is mainly volume based and not explicitly risk based.
- Goals of Solvency II: strengthen the capital adequacy by reducing the possibilities of consumer loss or market disruption in insurance (policyholder protection and financial stability motives)
- Solvency II is also based on a three-pillar system
  Pillar 1 quantification of regulatory capital
  Pillar 2 governance and supervision
  Pillar 3 disclosure of information to the public
- Under Pillar 1, a company calculates its solvency capital requirement (SCR) = minimal amount of capital ensuring that the probability of insolvency over a one-year period is no more than 0.5%. If this level of capital is not reached it will likely result in regulatory intervention and require remedial action.
- The firm must also calculate minimum capital requirement (MCR) = minimum capital to cover it's risks. If an insurer violates the MCR constraint, it will be prohibited from writing any further business.
- For calculating capital requirements, a standard formula or an internal model may be used. Either way, a total balance sheet approach is taken (all risks and their interactions are considered) The insurer should have own funds (surplus of assets over liabilities) exceeding both SCR and MCR
- Under Pillar 2, the company must demonstrate that it has a RM system in place and that this system is integrated into decision making processes
- An internal model must pass the "use test": It must be an integral part of the RM system and be actively used in the running of the firm. Moreover, a firm must undertake an ORSA (own risk and solvency assessment)
- An internal model often takes the form of a so-called economic scenario generator (ESG) in which risk-factor scenarios for a one-year period are randomly generated and applied to determine the SCR.

ORSA (Own risk and solvency assessment)

ORSA = Entirety of processes and procedures to identify, assess, monitor, manage, and report short and long term risks a (re)insurance company may face and to determine the own funds necessary to ensure the company's solvency at all times.
ORSA (Pillar 2) is different from capital calculations (Pillar 1).

- ORSA refers to a process (and not just an exercise in regulatory compliance)
- Each firm's ORSA is its own process and likely to be unique (not bound by a common set of rules such as the standard-formula approach in Pillar 1)
- ORSA goes beyond the one-year time horizon (which is a limitation of Pillar 1); e.g. for life insurance

Benefits & Criticism of regulatory frameworks

Benefits of regulation: Customer protection, responsible corporate governance, fair and comparable accounting rules, transparent information on risk, capital and solvency for shareholders etc.
Criticism:

- Costs and complexity for setting up and maintaining a sound risk management system compliant with present regulations (in the UK, Solvency II compliance costs at least 3 billion pounds). Regulation becomes more and more complex.
- Endogenous risk Regulation may amplify shocks. It can lead to risk-management herding (institutions all run for the same exit by following the same (perhaps VaR-based) rules in times of crisis and thus further destabilize the whole system).
- Market consistent valuation (at the core of the Basel rules for the trading book and Solvency II) implies that capital requirements are closely coupled to volatile financial markets.

Why manage financial risk?

Societal view

- Society relies on the stability of the banking and insurance system. The regulatory process, from which Basel II and Solvency II resulted, was motivated by the desire to prevent insolvency of individual institutions and thus protect customers (microprudential perspective)
- Since the 2007–2009 crisis, the reduction of systemic risk has become an important secondary focus (macroprudential perspective)
- The interests of society are served by enforcing the discipline of risk management in financial firms, through the use of regulation. 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. Individual firms need to be allowed to fail on occasion.

Shareholders' view

- While individual investors are typically risk averse and should therefore manage the risk in their portfolios, it is not immediately clear that risk management at the corporate level (e.g. holding a certain amount of risk capital or hedging a foreign currency exposure) increases the value of a corporation and thus enhances shareholder value.
- Theoretically, if investors have access to perfect capital markets, they can incorporate RM via their own trading and diversification.
- The Modigliani–Miller theorem, which marks the beginning of modern corporate finance theory, states that, in an ideal world without taxes, bankruptcy costs and informational asymmetries, and with frictionless and arbitrage-free capital markets, the financial structure of a firm (thus its RM decisions) is irrelevant for a firm's value.

Reasons for corporate RM

- RM can reduce taxes.
- RM can be beneficial, since a company may have better access to capital markets than individual investors.
- RM can increase the firm value in the presence of bankruptcy costs (liquidation costs or litigation costs), as it reduces the likelihood of bankruptcy.
- RM can reduce the impact of costly external financing.

2 Basic Concepts in Risk Management

2.1 Risk management for a financial firm

Empirical Distribution Function

Let $x_1, \ldots, x_n$ be independent realizations of a random variable X. The corresponding empirical distribution function $\hat{F}_X : \mathbb{R} \to [0,1]$ is given by the step function

$$\hat{F}_X(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \le x\}}, \quad x \in \mathbb{R},$$

where $1_{\{.\}}$ is the indicator function.

Asset, liabilities and the balance sheet

Balance sheet eq: Assets = Liabilities + 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 at book value at its inception and then updates it over time.
- Fair value accounting: Tries to value assets and liabilities at market prices. This can be challenging for illiquid assets or liabilities.

There is a tendency in the industry to move towards fair value accounting.

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 defaults of loans or securities trading)
- Dry up of funding liquidity Rates for short-term funding can increase suddenly.
- Maturity mismatch (especially for banks, 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 insurance company 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 exposed to different sources of risk (e.g. interest rate risk or longevity risk).
- So risk is found on both sides of the balance sheet and thus RM should not focus on the asset side alone

2.2 Modeling value and value change

Different notions of capital

- Equity capital
  - Value of assets - liabilities
  - Measures the firm's value to its shareholders
  - Contains initial capital invested in the firm and retained earnings (accumulated earnings not paid out to shareholders).
- Regulatory capital
  - Capital required according to regulatory rules
  - For European insurance firms: SCR and MCR
  - A regulatory framework also specifies the capital quality. One distinguishes Tier 1 capital (best quality of capital such as retained earnings, common stock, non-redeemable preferred stock), Tier 2 capital (lower quality of capital such as undisclosed reserves, revaluation reserves, hybrid instruments and subordinated term debt) and Tier 3 capital (tertiary capital; e.g. undisclosed reserves and debt of lower quality than in tier 2)
- 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 to work with fair values of balance sheet items.

Risk Mappings

- Consider a portfolio of assets and liabilities with time-t value V_t
- Δt is a time increment (used as time unit); For small Δt, we usually assume,
  - the portfolio composition remains unchanged over Δt
  - there are no intermediate payments during Δt
- Value change: $\Delta V_{t+1} = V_{t+1} - V_t$
- One-period-ahead loss: $L_{t+1} = -\Delta V_{t+1}$
- The loss distribution is the distribution of $L = L_{t+1}$; that is, the measure μ_L on $\mathbb{R}$ given by $\mu_L = L^{\#}\mathbb{P} = \mathbb{P} \circ L^{-1}$ (push-forward)
- Loss distribution is fully specified by the cdf $F_L : \mathbb{R} \to [0,1]$, $F_L(x) = \mathbb{P}[L \leq x]$. It satisfies
  (1) Normalization $\lim_{x\to-\infty} F_L(x) = 0, \lim_{x\to\infty} F_L(x) = 1$
  (2) Right-continuity $F_L(x_n) \downarrow F_L(x)$ for $x_n \downarrow x \in \mathbb{R}$
  (3) Monotonicity $F_L(a) \leq F_L(b)$ for $a \leq b$
- Carathéodory's extension theorem: Every func $F : \mathbb{R} \to [0,1]$ satisfying (1)-(3) is a cdf of a RV L
- If F_L is absolutely continuous wrt the Lebesgue measure, there exists a measurable func $f_L : \mathbb{R} \to \mathbb{R}_+$ s.t. $F_L(x) = \int_{-\infty}^{x} f_L(y)dy$, f_L is called pdf, or simply density, of L.
- Often consider the profit-and-loss (P&L) dist, which is the dist of $\Delta V_{t+1} = -L_{t+1}$.
- For longer time intervals, one sometimes considers the discounted P&L $\Delta V_{t+1} = \frac{V_{t+1}}{1+r} - V_t$, where r is risk-free interest rate.
- V_t is often modeled as a function of time and a vector $\mathbf{Z}_t = \left(Z_t^1, \ldots, Z_t^d\right)$ of risk factors $V_t = f(t, \mathbf{Z}_t)$ for some measurable mapping $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$
- Changes in risk factors: $\mathbf{X}_{t+1} = \mathbf{Z}_{t+1} - \mathbf{Z}_t$
- L_{t+1} can be written in terms of L_t and X_{t+1} as
$$\begin{aligned} L_{t+1} &= -V_{t+1} + V_t = -f(t+1, \mathbf{Z}_{t+1}) + f(t, \mathbf{Z}_t) \\ &= -f(t+1, \mathbf{Z}_t + \mathbf{X}_{t+1}) + f(t, \mathbf{Z}_t) \end{aligned}$$
- We differentiate between the unconditional distribution of L_{t+1} and its conditional distribution given $\mathcal{F}_t$, describing the information available at time t. The two cdf's are $\mathbb{P}[L_{t+1} \leq x]$ and $\mathbb{P}[L_{t+1} \leq x \mid \mathcal{F}_t]$
- Usually, $\mathbf{Z}_t$ is assumed to be known at time t.

Assume $\mathbf{Z}_t$ is known and $\mathbf{X}_{t+1}$ is random

- If f is differentiable, one obtains by first-order Taylor approximation, $f(t+1, \mathbf{Z}_t + \mathbf{X}_{t+1}) \approx f(t, \mathbf{Z}_t) + f_t(t, \mathbf{Z}_t) + \sum_{j=1}^{d} f_{z^j}(t, \mathbf{Z}_t) X_{t+1}^j$
- Then L_{t+1} can be approximated by the linearized loss $L_{t+1}^{\Delta} = -f_t(t, \mathbf{Z}_t) - \sum_{j=1}^{d} f_{z^j}(t, \mathbf{Z}_t) X_{t+1}^j$

  For given Z_t, it is a linear func of $X_{t+1}^1, \ldots, X_{t+1}^d$: $L_{t+1}^{\Delta} = -c_t - \mathbf{b}_t^{\top} \mathbf{X}_{t+1}$
- The approx is best for small risk-factor changes.

Stock Portfolio

- Consider d stocks with time-t values $S_t^1, \ldots, S_t^d$
- Numbers of stocks held at time $t : \lambda^1, \ldots, \lambda^d$
- Portfolio value: $V_t = \sum_{j=1}^{d} \lambda^j S_t^j = \sum_{j=1}^{d} \lambda^j e^{Z_t^j}$ for the log-prices $Z_t^j = \log S_t^j$
- One-period-ahead loss:
$$\begin{aligned} L_{t+1} &= -\sum_{j=1}^{d} \lambda^j \left(e^{Z_t^j + X_{t+1}^j} - e^{Z_t^j}\right) \\ &= -\sum_{j=1}^{d} \lambda^j S_t^j \left(e^{X_{t+1}^j} - 1\right) = -\sum_{j=1}^{d} w_t^j \left(e^{X_{t+1}^j} - 1\right) \end{aligned}$$
- Linear approximation: $e^{X_{t+1}^t} - 1 \approx X_{t+1}^j$
- Linearized loss: $L_{t+1}^{\Delta} = -\sum_{j=1}^{d} w_t^j X_{t+1}^j = -\mathbf{w}_t^{\top} \mathbf{X}_{t+1}$
- Cond mean vec: $\boldsymbol{\mu} := \mathbb{E}_t \mathbf{X}_{t+1} = \mathbb{E}[\mathbf{X}_{t+1} \mid \mathcal{F}_t]$
- Conditional Cov Matrix: $\Sigma_t := \mathrm{Cov}_t \mathbf{X}_{t+1}$ where $\mathrm{Cov}_t\left(X_{t+1}^i, X_{t+1}^j\right) = \mathbb{E}_t\left[\left(X_{t+1}^i - \mathbb{E}_t X_{t+1}^i\right)\left(X_{t+1}^j - \mathbb{E}_t X_{t+1}^j\right)\right]$
- Then the conditional expectation and conditional variance of L_{t+1}^{Δ} are
$$\mathbb{E}_t L_{t+1}^{\Delta} = -\mathbb{E}_t\left(\mathbf{w}_t^T \mathbf{X}_{t+1}\right) = -\mathbf{w}_t^T \boldsymbol{\mu}_t$$
$$\mathrm{Var}_t\left(L_{t+1}^{\Delta}\right) = \mathrm{Var}_t\left(\mathbf{w}_t^T \mathbf{X}_{t+1}\right) = \mathbf{w}_t^T \Sigma_t \mathbf{w}_t$$

European Call Option

- Consider a European call on a non-dividend paying stock S_t w/ maturity T and strike price K.
- The Black-Scholes price of the option is $C^{BS}(t, S_t, r, \sigma, K, T) = S_t \Phi(d_1) - K e^{-r(T-t)} \Phi(d_2)$
  - t is time in years
  - Φ is the standard normal cdf
  - r is continuously compounded risk-free interest
  - σ is the annualized volatility of S
  - $d_1 = \frac{\log(S_t/K) + (r+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$
  - $d_2 = d_1 - \sigma\sqrt{T-t}$
- In the Black-Scholes model, it is assumed that interest rates and volatilities are constant. But in reality they tend to fluctuate over time.
- We therefore add them to our vector of risk factors $\mathbf{Z}_t = (\log S_t, r_t, \sigma_t)$ and obtain the corresponding vector of risk-factor changes $\mathbf{X}_{t+1} = (\log(S_{t+1}/S_t), r_{t+1} - r_t, \sigma_{t+1} - \sigma_t)$
- If we are interested in daily losses, we measure time in units of $\Delta t = 1/250$ (250 is apx num of business days of one year). Then
  - $V_t = f(t, \mathbf{Z}_t) = C^{BS}(t\Delta t, S_t, r_t, \sigma_t, K, T)$
  - $f_t(t, \mathbf{Z}_t) = C_t^{BS}(t\Delta t, S_t, r_t, \sigma_t, K, T)\Delta t$
- Linearized loss
$$\begin{aligned} &L_{t+1}^{\Delta} \\ &= -f_t(t, \mathbf{Z}_t) - \sum_{j=1}^{3} f_{z^j}(t, \mathbf{Z}_t) X_{t+1}^j \\ &= -\left(C_t^{BS}\Delta t + C_S^{BS} S_t X_{t+1}^1 + C_r^{BS} X_{t+1}^2 + C_\sigma^{BS} X_{t+1}^3\right) \end{aligned}$$
  C_t^{BS} = theta, C_S^{BS} = delta, C_r^{BS} = rho, C_σ^{BS} = vega
- For portfolios of derivatives, the linear approximation L_{t+1}^{Δ} is not always good
- Higher order Taylor approximations can be used. E.g. delta-gamma approximation (second order approximation)

Valuation Methods

Fair Value Accounting

1. Mark-to-market: The fair value of an investment is determined from quoted prices for the same instrument.
2. Mark-to-model with objective inputs: The fair value of an instrument is determined using quoted prices in active markets for similar instruments or by using valuation techniques/models with inputs based on observable market data.
3. Mark-to-model with subjective inputs: The fair value of an instrument is determined using valuation techniques/models for which some inputs are not observable in the market (e.g. default risk of portfolios of loans to companies for which no CDS spreads are available)

Risk-neutral Valuation

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

European Call Option - Continued

- Suppose that options with a particular strike K^* and maturity T^* are not traded, but options with different strikes and maturities on the same stock are.
- Under $\mathbb{P}$, the stock price (S_t) is assumed to follow a geometric Brownian motion (GBM) (the so-called Black-Scholes model) with dynamics
$$dS_t = \mu S_t dt + \sigma S_t dW_t$$
$$\leftrightarrow \quad S_t = S_0 \exp\left(\mu t + \sigma W_t - \frac{1}{2}\sigma^2 t\right)$$
for constants $\mu \in \mathbb{R}$ (drift), $\sigma > 0$ (volatility) and and a standard Brownian motion (W_t).
- The model is complete $\Leftrightarrow$ there exists a unique EMM $\mathbb{Q}$
- Under $\mathbb{Q}, e^{-rt}S_t$ is a martingale and S_t follows a GBM with drift r and volatility σ
- The European call option payoff is $H = (S_T-K)^+ = \max\{S_T-K, 0\}$ and the risk-neutral valuation formula can be shown to be $\mathbb{E}_t^{\mathbb{Q}} e^{-r(T-t)}(S_T-K)^+ = C^{BS}(t, S_t, r, \sigma, K, T)$
- The risk-less interest rate r is given by the market. But it is difficult to predict the volatility σ
- One typically uses quoted prices $C^{BS}(t, S_t, r, \sigma, K, T)$ for different K and T to infer the σ used to price the option corresponding to K^* and T^* (implied volatility)

Different ways of generating loss distributions

1. Analytical method

Model f and $\mathbf{X}_t, t = 0, 1, \ldots$ such that the conditional distribution of the loss L_{t+1} or linearized loss L_{t+1}^{Δ} given $\mathcal{F}_t$ can be derived in closed form.

Example: f is differentiable and $\mathbf{X}_t, t = 0, 1, \ldots$ are iid $\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ Then, for given $\mathbf{Z}_t$,

$$L_{t+1}^{\Delta} = -f_t(t, \mathbf{Z}_t) - \sum_{j=1}^{d} f_{z_j}(t, \mathbf{X}_t) X_{t+1}^j$$
$$= -c_t - \mathbf{b}_t^\top \mathbf{X}_{t+1} \sim \mathcal{N}\left(-c_t - \mathbf{b}_t^\top \boldsymbol{\mu}, \mathbf{b}_t^\top \Sigma \mathbf{b}_t\right)$$

Advantage: easy to implement
Drawback: Normal distributions are not always a good description of financial data. $\mathbf{X}_{t+1}$ is often asymmetric and leptokurtic (the pdf has a thinner body, heavier tails than a normal pdf)

2. Historical simulation

Approximate the distribution of L_{t+1} by the edf (empirical distribution function) $\hat{F}_L(x) = \frac{1}{n}\sum_{i=1}^n 1_{\{l_{t-i+1} \le x\}}$, where $l_{t-n+1}, \ldots, l_t$ are the last n realized losses.
Advantages: easy to implement, no modeling assumptions, no estimation required
Drawbacks: Sufficient data for all risk-factors required, makes predictions based on past data.

3. Monte Carlo simulation

Build a model for L_{t+1} and simulate from it.
Use simulations $l_1, \ldots, l_n$ to generate the simulated distribution function $\hat{F}_L(x) = \frac{1}{n}\sum_{i=1}^n 1_{\{l_i \le x\}}$
Advantages: quite general, doesn't need many assumptions
Drawbacks: One needs a good model for L_{t+1}. It can be computationally costly to simulate a high-dimensional vector of risk factors.
So called economic scenario generators (i.e. economically motivated dynamic models for the evolution and interaction of risk factors) used in insurance also fall into the category of Monte Carlo methods.

2.3 Risk measurement

Risk measurement

- A risk measure assigns to a random loss L a real number $\rho(L)$ measuring the riskiness of L
- In a regulatory framework, the number is often interpreted as the amount of buffer capital needed to compensate for the risk of L
- Some reasons for using risk measures are:
  - 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 of a trader's position should not exceed a given bound).
  - To determine the riskiness (and thus fair premium) of an insurance contract.

Different approaches to risk measurement

1. Notional-amount approach

- risk of a portfolio = sum of the notional values of the securities times their riskiness factor
- oldest approach
- standardized approaches of Basel II (e.g. OpRisk) still use it
- Advantages: simplicity
- Drawbacks:
  - No differentiation between long and short positions, and no netting: the risk of a long position in corporate bonds hedged by an offsetting position in credit default swaps is counted as twice the risk of the unhedged bond position.
  - No diversification benefits: risk of a portfolio of loans to many companies = risk of a portfolio where the whole amount is lent to a single company.
  - Problems for portfolios of derivatives: notional amount of the underlying can widely differ from the economic value of the derivative position.

2. 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)
- The stressed loss corresponding to a collection $x_1, \ldots, x_n$ of risk factor changes with corresponding weights $w_1, \ldots, w_n$ is $\max\limits_{1\le i\le n} w_i L(x_i)$
- Mathematical Interpretation
  - Assume $L(0) = 0$ (OK for small Δt) and $w_i \in (0,1]$
  - $w_i L(x_i) = w_i L(x_i) + (1-w_i)L(0) = \mathbb{E}^{\mathbb{P}_i} L(\mathbf{X})$, where $\mathbf{X} \sim \mathbb{P}_i = w_i\delta_{x_i} + (1-w_i)\delta_0$ (δ_x Dirac measure at x) is a probability measure on $\mathbb{R}^d$
  - Then $\max\limits_{1\le i\le n} w_i L(x_i)$ $= \max\left\{\mathbb{E}^{\mathbb{P}} L(\mathbf{X}) : \mathbf{X} \sim \mathbb{P} \in \{\mathbb{P}_1, \ldots, \mathbb{P}_n\}\right\}$ can be seen as a worst case expected loss (related to coherent risk measures)
- Advantages: Easy to implement; Useful complementary information to risk measures based on loss distributions.
- Drawbacks: how does one determine the scenarios and the weights?

3. Risk measures based on loss distributions

Many modern risk measures are characteristics of the underlying (conditional or unconditional) loss dist over some predetermined time horizon Δt
Advantage:

- The concept of a loss distribution makes sense on all levels (from single portfolios to the overall position of a financial institution).
- If estimated properly, loss distributions reflect netting and diversification effects.

Drawbacks:

- Estimates of loss distributions are typically based on past data
- It is difficult to estimate loss distributions accurately (especially for large portfolios). $\rightsquigarrow$ Risk measures should be complemented by information from scenarios (forward-looking)

Quantiles

Let L be a RV with cdf $F_L(x) = \mathbb{P}[L \le x]$ (non-decreasing and right-continuous). Let $\alpha \in (0,1)$.
(1) α-quantile is any $q \in \mathbb{R}$ s.t. $\mathbb{P}[L < q] \le \alpha \le \mathbb{P}[L \le q]$
(2) Left- α-quantile: $q_L^-(\alpha) = \sup\{x \in \mathbb{R} : F_L(x) < \alpha\} = \min\{x \in \mathbb{R} : F_L(x) \ge \alpha\}$
(3) Right- α-quantile: $q_L^+(\alpha) = \sup\{x \in \mathbb{R} : F_L(x) \le \alpha\} = \inf\{x \in \mathbb{R} : F_L(x) > \alpha\}$

- $q_L^-(\alpha)$ and $q_L^+(\alpha)$ are both α-quantiles
- $q_L^-(\alpha)$ is non-decreasing and left-continuous in α
- $q_L^+(\alpha)$ is non-decreasing & right-continuous in α

Let $\alpha \mapsto q(\alpha)$ be an arbitrary quantile-function of L, and fix $x \in \mathbb{R}$.

- Then $\alpha < F_L(x) \Rightarrow q(\alpha) \le x, q(\alpha) \le x \Rightarrow \alpha \le F_L(x)$
- So $\{\alpha \in (0,1) : \alpha < F_L(x)\} \subseteq \{\alpha \in (0,1) : q(\alpha) \le x\} \subseteq \{\alpha \in (0,1) : \alpha \le F_L(x)\}$ and
- $\eta[\alpha \in (0,1) : q(\alpha) \le x] = F_L(x)$, where η is the Lebesgue measure on $(0,1)$

This shows that $\alpha \mapsto q(\alpha)$ has the same distribution under η as L under $\mathbb{P}$.

Value-at-Risk and Expected Shortfall

Value-at-Risk at level α: $\mathrm{VaR}_\alpha(L) = q_L^-(\alpha) = \min\{x \in \mathbb{R} : \mathbb{P}[L - x \le 0] \ge \alpha\}$
Expected Shortfall at level α: $\mathrm{ES}_\alpha(L) = \mathbb{E}[L \mid L \ge \mathrm{VaR}_\alpha(L)]$

- VaR_α is defined on $L^0(\mathbb{P})$ (all random variables)
- $\mathrm{VaR}_\alpha(L)$ is non-decreasing in α
- $\mathrm{VaR}_\alpha(L) \ge \mathrm{VaR}_\alpha(L')$ if $L \ge L'$ $\mathbb{P}$-almost surely, that is, $\mathbb{P}[L \ge L'] = \mathbb{P}[\{\omega \in \Omega : L(\omega) \ge L'(\omega)\}] = 1$
- Linearity: $\mathrm{VaR}_\alpha(a + bL) = a + b\,\mathrm{VaR}_\alpha(L)$ for $a \in \mathbb{R}$ and $b \in (0,\infty)$
- ES_α is defined on $L^1(\mathbb{P})$
- $\mathrm{ES}_\alpha(L)$ is non-decreasing in α
- Linearity: $\mathrm{ES}_\alpha(a + bL) = a + b\mathrm{ES}_\alpha(L)$ for $a \in \mathbb{R}$ and $b \in (0,\infty)$
- $\mathrm{VaR}_\alpha(L) \le \mathrm{ES}_\alpha(L)$
- VaR_α does not see what happens in the tail (frequency measure)
- ES_α looks into the tail (severity measure), but is more difficult to estimate and backtest than VaR_α (larger sample size required)
- VaR may give incentives to concentrate risk!

4. Coherent risk measures

Let $\mathcal{L}$ be a vector space of random variables L on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (modeling uncertain losses); e.g. $\mathcal{L} = L^p(\mathbb{P})$ for some $p \in \{0\} \cup [1, \infty]$.
A mapping $\rho : \mathcal{L} \to \mathbb{R}$ is a **coherent risk measure** if it satifies
(M) Monotonicity: $\rho(L_1) \le \rho(L_2)$ for $L_1 \le L_2$ $\mathbb{P}$-almost surely (i.e. $\mathbb{P}[L_1 \le L_2] = 1$)
(T) Translation property: $\rho(L+m) = \rho(L)+m$ for $m \in \mathbb{R}$
(S) Subadditivity: $\rho(L_1 + L_2) \le \rho(L_1) + \rho(L_2)$
(P) Positive homogeneity: $\rho(\lambda L) = \lambda\rho(L)$ for $\lambda \in \mathbb{R}_+$
and a **convex risk measure** if it satisfies (M), (T) together with
(C) Convexity: $\rho(\lambda L_1 + (1-\lambda)L_2) \le \lambda\rho(L_1) + (1-\lambda)\rho(L_2)$ for all $0 < \lambda < 1$
Note that under (P), one has (S) $\Leftrightarrow$ (C). So every **coherent risk measure is also convex**.
Sometimes risk measures are defined for the **P&L** $-L$ instead of the **loss** L.
VaR_α and ES_α are distribution-based (only depend on the distribution of L under $\mathbb{P}$).
VaR_α is distribution-based and satisfies (M), (T), (P). On the other hand, it does not satisfy (S), and as a consequence, also not (C)

Worst Expected Losses

Let $\mathcal{P}$ be an arbitrary non-empty set of probability measures $\mathbb{Q} \ll \mathbb{P}$. Then $\rho_{\mathcal{P}}(L) = \sup_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}^{\mathbb{Q}} L$ is a coherent risk measure.
Can be regard as a generalized form of the **stress test risk measures**.
Note that, in general, $\rho_{\mathcal{P}}$ is not a functional of the distribution of L under $\mathbb{P}$.

- VaR_α and ES_α are distribution-based (only depend on the distribution of L under $\mathbb{P}$
- VaR_α satisfies (M), (T), (P) but does not satisfy (S), and as a consequence, also not (C).

Average-Value-at-Risk

- For $L \in L^1(\mathbb{P})$, define $\mathrm{AVaR}_\alpha(L) := \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)du = \frac{1}{1-\alpha}\int_\alpha^1 q_L^-(u)du$.
- Like VaR_α and ES_α, AVaR_α is **distribution-based** and satisfies (M), (T), (P).
- If $\mathbb{P}\left[L \ge q_L^-(\alpha)\right] = 1-\alpha$ (in particular, if F_L is continuous), then $\mathrm{AVaR}_\alpha(L) = \mathrm{ES}_\alpha(L)$.
  Proof: Under the Lebesgue measure η, q_L^- is distributed like L under $\mathbb{P}$ If $\mathbb{P}\left[L \ge q_L^-(\alpha)\right] = 1-\alpha$, then, since ES_α is distribution-based,

$$\mathrm{ES}_\alpha[L] = \mathrm{ES}_\alpha^\eta\left[q_L^-\right] = \frac{1}{1-\alpha}\int_\alpha^1 q_L^-(u)du$$

Thm: Let $L \in L^1(\mathbb{P})$ and $\alpha \in (0,1)$. Then

$$\mathrm{AVaR}_\alpha(L) = \min_{s\in\mathbb{R}}\left(\frac{\mathbb{E}(L-s)^+}{1-\alpha} + s\right) \tag{1}$$

$$= \max_{\mathbb{Q}\in\mathcal{P}_\alpha} \mathbb{E}^{\mathbb{Q}} L \quad \text{for} \quad \mathcal{P}_\alpha = \left\{\mathbb{Q} : \left\|\frac{d\mathbb{Q}}{d\mathbb{P}}\right\|_\infty \le \frac{1}{1-\alpha}\right\} \tag{2}$$

Moreover, the min in (1) is attained by any α-quantile q of L, and the max in (2) by any $\mathbb{Q}$ with

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{1}{1-\alpha}1_{\{L>q\}} + \kappa 1_{\{L=q\}}.$$

for an α-quantile q and $\kappa \ge 0$ so that $\mathbb{Q}$ is a probability measure.

- The right side of (1) is called **CVaR (conditional value at risk)**
- (2) is a **worst expected loss** (or robust representation). It shows that AVaR_α is **coherent**.

Proof: Any quantile function q of L has the same distribution under η as L under $\mathbb{P}$. So

$$\mathrm{AVaR}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 q(u)du$$
$$= \frac{1}{1-\alpha}\int_\alpha^1 (q(u) - q(\alpha))du + q(\alpha)$$
$$= \frac{\mathbb{E}^\eta(q-q(\alpha))^+}{1-\alpha} + q(\alpha) = \frac{\mathbb{E}^{\mathbb{P}}(L-q(\alpha))^+}{1-\alpha} + q(\alpha)$$

For $\mathbb{Q}$ with $\frac{d\mathbb{Q}}{d\mathbb{P}} \le \frac{1}{1-\alpha}$ and $s \in \mathbb{R}$

$$\mathbb{E}^{\mathbb{Q}} L = \mathbb{E}^{\mathbb{Q}}(L-s) + s \le \frac{\mathbb{E}^{\mathbb{P}}(L-s)^+}{1-\alpha} + s$$

On the other hand, for any α-quantile q_α, if

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{1}{1-\alpha}1_{\{L>q_\alpha\}} + \kappa 1_{\{L=q_\alpha\}},$$

then $\mathbb{E}^{\mathbb{Q}} L = \mathbb{E}^{\mathbb{Q}}(L-q_\alpha) + q_\alpha = \frac{\mathbb{E}^{\mathbb{P}}(L-q_\alpha)^+}{1-\alpha} + q_\alpha$

ES and AVaR are almost the same

Let $L \in L^1(\mathbb{P})$ and $\alpha \in (0,1)$. We defined

1. **Expected Shortfall**: $\mathrm{ES}_\alpha(L) := \mathbb{E}[L \mid L \ge \mathrm{VaR}_\alpha(L)]$
2. **Average-VaR**: $\mathrm{AVaR}_\alpha(L) := \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)du$

- $\mathrm{ES}_\alpha(L) \le \mathrm{AVaR}_\alpha(L)$
- $\mathrm{ES}_\alpha(L) = \mathrm{AVaR}_\alpha(L)$ **if L has a continuous cdf**

Expected Shortfall, Average VaR, Conditional VaR or Tail VaR are the same thing.

VaR and AVaR for comonotonic random variables

Two RVs L_1 and L_2 on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are **comonotonic** if $L_1 = l_1(Z)$ and $L_2 = l_2(Z)$ for non-decreasing functions l_1, l_2 and a random variable Z (**perfect dependence**)
If L_1 and L_2 are comonotonic, then $\mathrm{VaR}_\alpha(L_1+L_2) = \mathrm{VaR}_\alpha(L_1) + \mathrm{VaR}_\alpha(L_2)$
Proof: L_i has the same distribution as $l_i\left(q_Z^-\right)$, and $L_1 + L_2$ has the same distribution as $l_1\left(q_Z^-\right) + l_2\left(q_Z^-\right)$
$\Rightarrow q_{L_i}^-$ is the left-continuous version of $l_i\left(q_Z^-\right)$, and $q_{L_1+L_2}^-$ is the left-continuous version of $l_1\left(q_Z^-\right) + l_2\left(q_Z^-\right) \Rightarrow q_{L_1+L_2}^-(\alpha) = q_{L_1}^-(\alpha) + q_{L_2}^-(\alpha)$ for all $\alpha \in (0,1)$.
and as a consequence, $\mathrm{AVaR}_\alpha(L_1+L_2) = \mathrm{AVaR}_\alpha(L_1) + \mathrm{AVaR}_\alpha(L_2)$
Proof:

$$\mathrm{AVaR}_\alpha(L_1+L_2) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L_1+L_2)du$$
$$= \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L_1)du + \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L_2)du$$
$$= \mathrm{AVaR}_\alpha(L_1) + \mathrm{AVaR}_\alpha(L_2)$$

Entropic Risk Measure

Let L be a RV and $\lambda > 0$ such that $\mathbb{E}\left[e^{\lambda L}\right] < \infty$

$$\mathrm{Ent}_\lambda(L) := \frac{1}{\lambda}\log\mathbb{E}\left[e^{\lambda L}\right]$$

- Clearly, Ent_λ is distribution-based and satisfies (M) and (T), but not (P)
- One has $\mathrm{Ent}_\lambda(L) = \max_{\mathbb{Q}\ll\mathbb{P}}\left(\mathbb{E}^{\mathbb{Q}}[L] - \frac{1}{\lambda}H(\mathbb{Q},\mathbb{P})\right)$, for the relative entropy $H(\mathbb{Q},\mathbb{P}) := \mathbb{E}^{\mathbb{Q}}\log\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)$, and the max is attained for $\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{e^{\lambda L}}{\mathbb{E}[e^{\lambda L}]}$
  This shows that Ent_λ is **convex**!

Proof: Define $\mathbb{Q}^L$ by $\frac{d\mathbb{Q}^L}{d\mathbb{P}} := \frac{e^{\lambda L}}{\mathbb{E}[e^{\lambda L}]}$.
Since $x \mapsto x\log x$ is strictly convex, one obtains from Jensen's inequality,

$$\mathbb{E}^{\mathbb{Q}}\log\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathbb{E}^{\mathbb{P}}\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\log\frac{d\mathbb{Q}}{d\mathbb{P}}\right)$$
$$\ge \mathbb{E}^{\mathbb{P}}\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\log\mathbb{E}^{\mathbb{P}}\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right) = 1 \times \log(1) = 0$$

for all $\mathbb{Q} \ll \mathbb{P}$ with equality iff $\mathbb{Q} = \mathbb{P}$. Therefore, $H(\mathbb{Q},\mathbb{P}) = \mathbb{E}^{\mathbb{Q}}\log\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathbb{E}^{\mathbb{Q}}\log\frac{d\mathbb{Q}}{d\mathbb{Q}^L} + \mathbb{E}^{\mathbb{Q}}\log\frac{d\mathbb{Q}^L}{d\mathbb{P}} \ge \mathbb{E}^{\mathbb{Q}}\log\frac{d\mathbb{Q}^L}{d\mathbb{P}} = \lambda\mathbb{E}^{\mathbb{Q}}[L] - \log\mathbb{E}^{\mathbb{P}}\left[e^{\lambda L}\right]$ with equality iff $\mathbb{Q} = \mathbb{Q}^L$. Hence,

$$\frac{1}{\lambda}\log\mathbb{E}^{\mathbb{P}}\left[e^{\lambda L}\right] = \max_{\mathbb{Q}\ll\mathbb{P}}\left(\mathbb{E}^{\mathbb{Q}}[L] - \frac{1}{\lambda}H(\mathbb{Q},\mathbb{P})\right)$$

3 Empirical Properties of Financial Data

3.1 Stylized facts of financial return series

- **Stylized facts** are a collection of empirical observations and inferences drawn of such, which apply to many time series of risk-factor changes (e.g. log-returns of equities, indices, exchange rates, commodity prices).
- Stylized facts often 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 sparsity; non-stationarity).
- Consider discrete-time risk-factor changes $X_t = Z_t - Z_{t-1}$, e.g. $Z_t = \log S_t$, in which 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 called **(log-)return**, the latter **(classical) return**.

Autocorrelation function (ACF)

$\rho_h = \mathrm{corr}(X_0, X_h)$ for $h \in \mathbb{Z}$
Non-zero ACF at lag 1 would imply a tendency for a return to be followed by a return of equal sign; Not the case $\rightsquigarrow$ sign of the next return cannot be predicted.
iid data would imply $\rho_X(h) = \rho_{|X|}(h) = 1_{\{h=0\}}$. Not the case $\rightsquigarrow$ (X_t) is not a random walk

Testing univariate distributions

- Let $x_1, \dots, x_n$ be realizations of iid random variables $X_1, \dots, X_n$ with cdf F
- **Empirical distribution function (edf)**

$$\hat{F}_n(x) := \frac{1}{n}\sum_{i=1}^n 1_{\{x_i \le x\}}$$

- **Law of Large Numbers**:

  For all $x \in \mathbb{R}$, $\quad \hat{F}_n(x) \to F(x) \quad$ $\mathbb{P}$-almost surely

- **Glivenko-Cantelli Theorem**:

$$\sup_{x\in\mathbb{R}}\left|\hat{F}_n(x) - F(x)\right| \to 0 \quad \mathbb{P}\text{-almost surely}$$

Standard statistical tests for general cdf F

For **general univariate** cdf F:

- **Kolmogorov-Smirnov test**: $T_n = \sup_{x\in\mathbb{R}}\left|\hat{F}_n(x) - F(x)\right|$
- **Cramér-von Mises test**: $T_n = n\int_{\mathbb{R}}\left[\hat{F}_n(x) - F(x)\right]^2 dF(x)$
- **Anderson-Darling test**:

$$T_n = n\int_{\mathbb{R}}\frac{\left[\hat{F}_n(x) - F(x)\right]^2}{F(x)(1-F(x))}dF(x)$$

For **normal** $F \sim \mathcal{N}\left(\mu, \sigma^2\right)$:

Jarque-Bera test: Let $\hat{\beta}_n$ and $\hat{\kappa}_n$ be sample versions of the (1) **skewness**: $\beta = \frac{\mathbb{E}(X-\mu)^3}{\sigma^3}$ and

(2) **kurtosis**: $\kappa = \frac{\mathbb{E}(X-\mu)^4}{\sigma^4}$

Then under the null-hypothesis, for large n, $\frac{n}{6}\left(\hat{\beta}_n^2 + \frac{(\hat{\kappa}_n - 3)^2}{4}\right) \sim \chi_2^2$

Note:

- Financial data is not symmetric, but neg-skewed.
- For normal dist, skewness = 0, kurtosis = 3

Graphical tests

Denote by $x_{(1)} \leq \ldots \leq x_{(n)}$ the **ordered sample** and note that

$$\hat{F}_n(x) := \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \leq x\}} = \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_{(i)} \leq x\}}$$

- **P-P plot**

$$\left(p_i, F\left(x_{(i)}\right)\right), \quad i = 1, \ldots, n,$$
$$\text{where} \quad p_i = \frac{i - 1/2}{n} \approx \frac{i}{n} \approx \hat{F}_n\left(x_{(i)}\right)$$

  If $\hat{F}_n \approx F$, the points are close to the diagonal.

- **Q-Q plot**

$$\left(q\left(p_i\right), x_{(i)}\right), \quad i = 1, \ldots, n,$$
$$\text{where} \quad u \mapsto q(u) \text{ is a quantile function of } F$$

  Again, if $\hat{F}_n \approx F$, the points are close to the diag.

- Typically, **tail differences are better visible in a Q-Q plot** than in a P-P plot. So, Q-Q plot are more widely used.

Q-Q Plot

- If $\hat{F}_n \approx N(0,1)$, points are close to the diagonal
- If $\hat{F}_n \approx N\left(\mu, \sigma^2\right)$, points are close to the line $y = \mu + \sigma x$
- S-shape hints at heavier tails than those of a normal distribution
- Daily returns typically have kurtosis $\kappa > 3$ **leptokurtic**: narrower center, heavier tails than $N\left(\mu, \sigma^2\right)$, for which $\kappa = 3$
- They typically have power-like tails

Longer time-interval return series

- By going from daily to weekly, monthly, quarterly and yearly data, these effects become less pronounced (returns look more iid, less heavy-tailed)
- The (**non-overlapping**) h-period log-return at $t \in \{h, 2h, \ldots, nh\}$ is

$$X_t^{(h)} = \log\left(\frac{S_t}{S_{t-h}}\right) = \log\left(\frac{S_t}{S_{t-1}} \frac{S_{t-1}}{S_{t-2}} \cdots \frac{S_{t-h+1}}{S_{t-h}}\right)$$
$$= \sum_{i=0}^{h-1} X_{t-i}$$

  A **Central Limit Theorem (CLT)** effect takes place (less heavy-tailed, less evidence of serial correlation)

- Problem: for larger h, less data is available
- Possible remedy: Consider **overlapping returns** $\left\{X_t^{(h)} : t \in \{h, h+k, h+2k, \ldots, h+nk\}\right\}$ for $1 \leq k < h$ $\rightsquigarrow$ more data, but **serially dependent** now

Summary: 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 tail)

3.2 Multivariate stylized facts

Correlation between different assets

Consider d-dimensional vectors of log-return data $X_1, \ldots, X_n$

- By (U1), the returns of stock A at times t and $t+h$ show little correlation
- The same applies to the returns of stock A at time t and stock B at time $t+h$
- Returns of two different stocks on the same day may be correlated due to common underlying factors (**contemporaneous dependence**)
- Contemporaneous correlations of returns vary over time (difficult to detect whether changes are continuous 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 $\rightsquigarrow$ Returns of large magnitude of stock A may be followed by returns of large magnitude of stocks A and B

Tail Dependence

- The normal distribution cannot replicate **tail dependence**
- The t_3 distribution can produce **joint large gains/losses** but in a **symmetric way**.

Summary: 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 time t)

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

(M3) Correlations between contemporaneous returns vary over time

(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

Basic Definitions

- A **time series** is a family of random variables $(X_t)_{t \in I}$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a non-empty index set $I \subseteq \mathbb{Z}$
- **Mean function**: $\mu(t) = \mathbb{E}[X_t]$, **Covariance function**: $\gamma(t,s) = \operatorname{Cov}(X_t, X_s)$
- $(X_t)_{t \in \mathbb{Z}}$ is **stationary** if for all $t, s \in \mathbb{Z}$ with $t \leq s$, $(X_t, \ldots, X_s)$ has the same distribution as $(X_{t+1}, \ldots, X_{s+1})$
- $(X_t)_{t \in \mathbb{Z}}$ is weakly stationary if $\mathbb{E}\left[X_t^2\right] < \infty$, $\mu(t) = \mu \in \mathbb{R}$ and $\gamma(t,s) = \gamma(t+1, s+1)$ for all $t, s \in \mathbb{Z}$

  In this case, $\gamma(t,s) = \gamma(0, |t-s|) = \gamma(|t-s|)$

- If $(X_t)_{t \in \mathbb{Z}}$ is stationary and $\mathbb{E}\left[X_t^2\right] < \infty$ for some t, then it is also weakly stationary
- The **(ACF) autocorrelation function** of a weakly stationary time series is given by $\rho(h) = \operatorname{corr}(X_0, X_h) = \frac{\gamma(h)}{\gamma(0)}$
- A weakly stationary time series (X_t) is **white noise** if $\mathbb{E}X_t = 0$ and $\operatorname{Cov}(X_t, X_s) = 0$ for $t \neq s$ **Notation** for white noise with $\operatorname{Var}(X_t) = \sigma^2$: $\quad \operatorname{WN}\left(0, \sigma^2\right)$
- If (X_t) is an iid sequence with $\mathbb{E}X_t = 0$ and $\operatorname{Var}(X_t) = \sigma^2$, we call it **strict white noise**.

  Notation: $\operatorname{SWN}\left(0, \sigma^2\right)$

4.2 (G)ARCH models for changing volatility

- **(G)ARCH** = **(generalized) autoregressive conditionally heteroscedastic**
- **Typical innovations**: $Z_t \sim \mathcal{N}(0,1)$ or $Z_t \sim t_\nu / \sqrt{\nu/(\nu-2)}$

ARCH

- Let $(Z_t)_{t \in \mathbb{Z}} \sim \operatorname{SWN}(0,1)$. $(X_t)_{t \in \mathbb{Z}}$ is an ARCH(p) process if it is stationary and satisfies

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k X_{t-k}^2$$

  where $\alpha_0 > 0$ and $\alpha_k \geq 0, k = 1, \ldots, p$

- Let $(\mathcal{F}_t)$ be filtration generated by ARCH(p) (X_t)
- $\operatorname{Var}(X_t \mid \mathcal{F}_{t-1}) = \mathbb{E}\left[\sigma_t^2 Z_t^2 \mid \mathcal{F}_{t-1}\right] = \sigma_t^2 \mathbb{E}\left[Z_t^2 \mid \mathcal{F}_{t-1}\right] = \sigma_t^2 \mathbb{E}\left[Z_t^2\right] = \sigma_t^2$
- **Volatility** σ_t (conditional standard deviation) is changing in time, depending on past values of the process
- ARCH models can thus capture volatility clustering (if one of $|X_{t-1}|, \ldots, \left|X_{t-p}\right|$ is large, X_t is drawn from a distribution with large variance)
- This is where "**autoregressive conditionally heteroscedastic**"comes from
- For an ARCH(1), if $\mathbb{E}\left(Z_t^4\right) < \infty$ and $\alpha_1 < \min\left\{1, \left(\mathbb{E}\left(Z_t^4\right)\right)^{-1/2}\right\}$, one can show that

$$\kappa(X_t) = \frac{\mathbb{E}\left(X_t^4\right)}{\left(\mathbb{E}\left(X_t^2\right)\right)^2} = \frac{\kappa(Z_t)\left(1 - \alpha_1^2\right)}{1 - \alpha_1^2 \kappa(Z_t)}$$

  So if $\kappa(Z_t) > 1$, then $\kappa(X_t) > \kappa(Z_t)$

  For Gaussian or t, $\kappa(X_t) > 3$ (**leptokurtic**)

GARCH

- Let $(Z_t)_{t \in \mathbb{Z}} \sim \operatorname{SWN}(0,1)$. $(X_t)_{t \in \mathbb{Z}}$ is a GARCH(p,q) process if it is 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 $\alpha_0 > 0$, $\alpha_k \geq 0, k = 1, \ldots, p$, $\beta_k \geq 0, k = 1, \ldots, q$

- If one of $|X_{t-1}|, \ldots, \left|X_{t-p}\right|$ or $\sigma_{t-1}, \ldots, \sigma_{t-q}$ is large, X_t is drawn from a distribution with (persistently) large variance. Periods of high volatility tend to be more persistent.
- If $\mathbb{E}\left(\alpha_1 Z_t^2 + \beta_1\right)^2 < 1$ or $(\alpha_1 + \beta_1)^2 < 1 - (\kappa(Z_t) - 1)\alpha_1^2$, one can show that

$$\kappa(X_t) = \frac{\kappa(Z_t)\left(1 - (\alpha_1 + \beta_1)^2\right)}{1 - (\alpha_1 + \beta_1)^2 - (\kappa(Z_t) - 1)\alpha_1^2}$$

  So if $\kappa(Z_t) > 1$ (Gaussian, scaled t innovations), then $\kappa(X_t) > \kappa(Z_t)$

Volatility forecasting and risk measure estimation with GARCH(1,1)

- Assume you have observed the losses $L_0, L_1, \ldots, L_t$
- Fit a GARCH(1,1) model to it:

$$L_{t+1} = \sigma_{t+1} Z_{t+1}$$
$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 L_t^2 + \beta_1 \sigma_t^2$$

with e.g. $Z_{t+1} \sim \mathcal{N}(0,1)$ or $Z_{t+1} \sim t_\nu / \sqrt{\nu/(\nu-2)}$

- $F_{L_{t+1}|\mathcal{F}_t}(x) = \mathbb{P}[\sigma_{t+1} Z_{t+1} \le x \mid \mathcal{F}_t] = F_{Z_{t+1}}\left(\frac{x}{\sigma_{t+1}}\right)$
- $\mathrm{VaR}_\alpha^t(L_{t+1}) = \sigma_{t+1} \mathrm{VaR}_\alpha(Z_{t+1})$ and $\mathrm{ES}_\alpha^t(L_{t+1}) = \sigma_{t+1} \mathrm{ES}_\alpha(Z_{t+1})$

5 Extreme Value Theory

5.1 Maxima

Limiting behavior of averages

- Consider losses given by independent random variables $X_1, X_2, \ldots$ with cdf F
- If $\mathbb{E}[X_1] = \mu$, the SLLN (strong law of large numbers) gives $\frac{1}{n}\sum_{i=1}^n X_i \to \mu$ a.s.
- If $\mathbb{E}[X_1] = \mu$ and $\mathrm{Var}(X_1) = \sigma^2 < \infty$, the CLT (central limit theorem) gives $\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i - \mu}{\sigma} \to \mathcal{N}(0,1)$ in distribution

Limiting behavior of maxima

- The block maximum of $X_1, \ldots, X_n$ is given by $M_n = \max\{X_1, \ldots, X_n\}$
- $F_{M_n}(x) = \mathbb{P}[M_n \le x] = \mathbb{P}[X_1 \le x, \ldots, X_n \le x] = F^n(x), \quad x \in \mathbb{R}$
- One can show that $M_n \to x_F$ a.s., where $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\} \le \infty$ denotes the right end point of F (similar to SLLN)

Slowly Varying

A func $L : [a, \infty) \to (0, \infty)$ is sv (slowly varying) if $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$ for all $t > 0$

Maximum Domain of Attraction (MDA)

If there exist normalizing sequences of real numbers $\{c_n\} > 0$ and $\{d_n\}$ such that $(M_n - d_n)/c_n$ converges in distribution; that is,

$$\mathbb{P}[(M_n - d_n)/c_n \le x] = F^n(c_n x + d_n) \to H(x)$$

for all continuity points x of a non-degenerate cdf H (not a unit jump), then F is said to be in the MDA of H. Notation: $F \in \mathrm{MDA}(H)$

- We say H is of the same type as $\tilde{H}$ if $H(x) = \tilde{H}(cx + d)$ for some $c > 0$ and $d \in \mathbb{R}$
- Convergence to types theorem: If $\mathbb{P}[(M_n - d_n)/c_n \le x] \to H(x)$ for all continuity points of H and $\mathbb{P}\left[\left(M_n - \tilde{d}_n\right)/\tilde{c}_n \le x\right] \to \tilde{H}(x)$ for all continuity points of $\tilde{H}$, then H and $\tilde{H}$ are of the same type

Reverse Weibull case: $\xi < 0$

$F \in \mathrm{MDA}\left(H_\xi\right) \Leftrightarrow x_F < \infty$ and $\overline{F}\left(x_F - \frac{1}{x}\right) = 1 - F\left(x_F - \frac{1}{x}\right) = x^{\frac{1}{\xi}} L(x), x > 0$ for some sv func L.

All $F \in \mathrm{MDA}\left(H_\xi\right)$ share $x_F < \infty$

Examples: Uniform and Beta

Gumbel case: $\xi = 0$

- Ccontains dists whose tails decay roughly exponentially (we call these dists light-tailed).
- All moments exist for dists in Gumbel class
- Examples: normal, log-normal, exp, gamma

Fréchet case: $\xi > 0$

- $F \in \mathrm{MDA}\left(H_\xi\right) \Leftrightarrow \overline{F}(x) = 1 - F(x) = x^{-1/\xi} L(x), x > 0$ for some sv function L
- Distributions in $\mathrm{MDA}\left(H_\xi\right)$ for $\xi > 0$ are those whose tails decay like power functions.
- Examples: Student- t, Pareto, log-gamma, inverse gamma, Cauchy, α-stable with $0 < \alpha < 2$
- Example - Pareto distribution: $\overline{F}(x) = \left(\frac{\kappa}{\kappa + x}\right)^\theta, \quad \theta, \kappa > 0, \quad x \ge 0$

$$\overline{F}(x) = \left(1 + \frac{x}{\kappa}\right)^{-\theta} = x^{-\theta} L(x) \quad \text{for } L(x) = \left(\frac{\kappa x}{\kappa + x}\right)^\theta$$

So F is in $\mathrm{MDA}(H_{1/\theta})$

Generalized Extreme Value (GEV) Distribution

The standard generalized extreme value (GEV) distribution is given by its cdf and pdf

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

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

Support is all s s.t. $1 + \xi s > 0$. ξ is the shape param.
Generalized extreme value (GEV) distribution with shape param $\xi \in \mathbb{R}$, location param $\mu \in \mathbb{R}$ and scale param $\sigma > 0$ is of the form $H_{\xi,\mu,\sigma}(x) = H_\xi((x - \mu)/\sigma)$.
The shape param ξ determines the tail of the dist:

- $\xi < 0$ gives a Reverse Weibull distribution: short-tailed, $x_{H_\xi} = -1/\xi < \infty$
- $\xi = 0$ gives a Gumbel distribution: $x_{H_\xi} = \infty, \overline{H}_\xi(x) = 1 - H_\xi(x) \approx e^{-x}$
- $\xi > 0$ gives a Fréchet distribution: $x_{H_\xi} = \infty$, heavy-tailed.
  $\overline{H}_\xi(x) = 1 - H_\xi(x) \approx (\xi x)^{-1/\xi}$.
  $\mathbb{E}\left[X^k\right] = \infty \Leftrightarrow k \ge 1/\xi$, (tail becomes heavier for larger ξ)

Thm (Fisher–Tippett): If $F \in MDA(H)$, then H is of the form $H_{\xi,\mu,\sigma}$

Maxima of stationary time series

- Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary time series with stationary distribution $X_t \sim F$
- Let $\tilde{X}_t \sim F$ be iid and $\tilde{M}_n = \max\left\{\tilde{X}_1, \ldots, \tilde{X}_n\right\}$
- For many processes one can show that there exists a real number $\theta \in (0,1]$ such that $\lim_{n\to\infty} \mathbb{P}[(M_n - d_n)/c_n \le x] = H^\theta(x) \Leftrightarrow \lim_{n\to\infty} \mathbb{P}\left[\left(\tilde{M}_n - d_n\right)/c_n \le x\right] = H(x)$ (non-degenerate), θ is known as the extremal index
- If $F \in \mathrm{MDA}\left(H_\xi\right)$ for some ξ, then $(M_n - d_n)/c_n$ converges in distribution to H_ξ^θ Since H_ξ^θ is of the same type as H_ξ, the limiting distribution of the block maxima of the dependent series is the same as in the iid case **(only location and scale may change)**
- For large n,

$$\mathbb{P}[(M_n - d_n)/c_n \le x] \approx H^\theta(x) \approx F^{\theta n}(c_n x + d_n).$$

So the distribution M_n of a time series with extremal index θ can be approximated by the distribution $\tilde{M}_{\theta n}$ of the maximum of $\theta n < n$ observations from the associated iid series.
$\Rightarrow \theta n$ counts the number of roughly independent clusters in n observations
θ is often interpreted as 1/mean cluster size

- If $\theta = 1$, large sample maxima behave as in the iid case;
  if $\theta \in (0,1)$, large sample maxima tend to cluster
- Examples: Strict white noise: $\theta = 1$, GARCH processes: $\theta \in (0,1)$

5.2 Block maxima method

- Assume n is so large that $\mathbb{P}[(M_n - d_n)/c_n \le x] \approx H_\xi(x)$ is a good approximation
- For $y = c_n x + d_n, \mathbb{P}[M_n \le y] \approx H_\xi((y - d_n)/c_n) = H_{\xi,d_n,c_n}(y)$
- We collect data on block maxima and fit the three-parameter form of the GEV; that is, we wish to estimate $\xi, \mu = d_n, \sigma = c_n$
- Assume we have block maxima data $M_n^1, \ldots, M_n^m$ from m blocks of size n
- We want to estimate $\theta = (\xi, \mu, \sigma)$
- We construct a log-likelihood by assuming we have independent observations from a GEV with density h_θ, given by

$$\ell\left(\theta; M_n^1, \ldots, M_n^m\right) = \sum_{i=1}^m \log\left(h_\theta\left(M_n^i\right)\right)$$

where $h_\theta(x) = h_\xi((x - \mu)/\sigma)/\sigma$ with

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

- This can be maximized with respect to θ to obtain the MLE: $\hat{\theta} = (\hat{\xi}, \hat{\mu}, \hat{\sigma})$

Remarks

The fitted GEV model can be used to estimate the (1) size of an event with prescribed frequency (return level problem); (2) frequency of an event with prescribed size (return period problem)

Definition (Return Level; Return Period)

(1) The indicator in the density h_θ makes life difficult. The support depends on the parameters and it introduces non-differentiability. This does not allow the classical MLE regularity conditions for consistency and asymptotic efficiency to be applied

(2) For $\xi > -1/2$ (fine for practice), it can be shown that the MLE is regular

(3) Sufficiently many/large blocks require large amounts of data

(4) In defining blocks, bias and variance must be traded off (bias-variance tradeoff)

- Bias is reduced by increasing the block size n
- Variance is reduced by increasing the number of blocks m
- There is no general best strategy known to find the optimal block size

Confidence intervals for $r_{n,k}, k_{n,u}$ can be constructed via profile-likelihoods.

Return levels and return periods

The kn-block return level is $r_{n,k} = q_{M_n}^{\leftarrow}(1 - 1/k) = \mathrm{VaR}_{1-1/k}(M_n)$
The return period of the event $\{M_n > u\}$ is $k_{n,u} = 1/\overline{F}_{M_n}(u) = 1/\left(1 - F_{M_n}(u)\right)$

- $\mathbb{P}\left[M_n > r_{n,k}\right] \approx 1/k$. So $r_{n,k}$ is the level which is expected to be exceeded in one out of every k blocks of size n. E.g. 10-year return level $r_{260,10}$ = level exceeded in one out of every 10 years $\approx$ 2600 days
- $r_{n,k_{n,u}} \approx u$. So $k_{n,u}$ is the number of n-blocks for which we expect to see single n-block exceeding u

We use the approximation $F_{M_n} \approx H_{\hat{\xi},\hat{\mu},\hat{\sigma}}$ Then, parametric estimators for $r_{n,k}$ and $k_{n,u}$ are given by

$$\hat{r}_{n,k} = H_{\hat{\xi},\hat{\mu},\hat{\sigma}}^{-1}(1 - 1/k)$$
$$= \begin{cases} \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left((-\log(1-1/k))^{-\hat{\xi}} - 1\right) & \text{if } \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma}\log(-\log(1-1/k)) & \text{if } \hat{\xi} = 0 \end{cases}$$
$$\hat{k}_{n,u} = 1/\overline{H}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(u)$$

5.3 Peaks over threshold

Generalized Pareto distribution (GPD)

GPD is given by its cdf and pdf

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

$$g_{\xi,\beta}(x)=\begin{cases}\frac{1}{\beta}(1+\xi x/\beta)^{-(1+1/\xi)}1_{\{x\geq 0\}} & \xi>0,\\ \frac{1}{\beta}(1+\xi x/\beta)^{-(1+1/\xi)}1_{\{0\leq x<-\beta/\xi\}} & \xi<0\\ \frac{1}{\beta}\exp(-x/\beta)1_{\{x\geq 0\}} & \xi=0\end{cases}$$

where $\beta>0$, and the support is $x\geq 0$ for $\xi\geq 0$ and $0\leq x\leq -\beta/\xi$ for $\xi<0$. ξ is shape; β is 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 $G_{\xi,\beta}$.

  For $\xi>0$: $\mathbb{E}\left[X^k\right]=\infty\Leftrightarrow k\geq 1/\xi$;

  For $\xi\leq 0$: $\mathbb{E}[X]=\beta/(1-\xi)$
- $G_{\xi,\beta}\in\mathrm{MDA}\left(H_\xi\right)$

Excess Distribution; Mean Excess Function

The excess distribution above the threshold u is given by

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

If $\mathbb{E}|X|<\infty$, the mean excess function is (by Fubini's Theorem)

$$e(u)=\mathbb{E}[X-u\mid X>u]=\frac{\mathbb{E}\left[\int_u^\infty 1_{\{X>x\}}dx\right]}{\bar{F}(u)}$$
$$=\frac{\int_u^\infty\mathbb{E}\left[1_{\{X>x\}}\right]dx}{\bar{F}(u)}=\frac{1}{\bar{F}(u)}\int_u^{x_F}\bar{F}(x)dx$$

- F_u describes the distribution of the excess loss over u, given that u is exceeded
- $e(u)$ is the mean of F_u.
- If $\mathbb{E}|X|<\infty$ and the cdf F is continuous, then $\mathrm{ES}_\alpha(X)=e(\mathrm{VaR}_\alpha(X))+\mathrm{VaR}_\alpha(X)$
- If $F=G_{\xi,\beta}$, $\xi\neq 0$, then F_u is GPD with the same shape ξ and scale $\beta+\xi u$ (grows linearly in u). If $\xi\in(-\infty,0)\cup(0,1)$, then $e(u)=\frac{\beta+\xi u}{1-\xi}$, again linear in u. This is a characterizing property of GPD.

Theorem (Pickands–Balkema–de Haan)

There exists a positive measurable function β such that

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

if and only if $F\in\mathrm{MDA}\left(H_\xi\right)$

Peaks over threshold (POT) method

- Consider losses $X_1,\ldots,X_n\sim F\in\mathrm{MDA}\left(H_\xi\right)$
- Let $N_u=\#\{i\in\{1,\ldots,n\}:X_i>u\}$ be the number of exceedances of $u,\tilde{X}_1,\ldots,\tilde{X}_{N_u}$
- If the excesses $Y_i=\tilde{X}_i-u, i=1,\ldots,N_u$ are iid and (roughly) distributed as $G_{\xi,\beta}$, the log-likelihood is given by

$$\ell\left(\xi,\beta;Y_1,\ldots,Y_{N_u}\right)=\sum_{i=1}^{N_u}\log g_{\xi,\beta}\left(Y_i\right)$$
$$=\begin{cases}-N_u\log(\beta)-(1+\frac{1}{\xi})\sum_{i=1}^{N_u}\log\left(1+\frac{\xi Y_i}{\beta}\right) & \xi\neq 0\\ -N_u\log(\beta)-\sum_{i=1}^{N_u}\frac{Y_i}{\beta} & \xi=0\end{cases}$$

$\Rightarrow$ maximize wrt ξ and β s.t. $1+\xi Y_i/\beta>0$ for $i=1,\ldots,N_u$

Excesses over higher thresholds

- Assume $F_u\sim G_{\xi,\beta}$, then $F_v\sim G_{\xi,\beta+\xi(v-u)}$
- The excess distribution over $v\geq u$ remains GPD with the same ξ, and β grows linearly in v
- If $\xi\in(-\infty,0)\cup(0,1)$, the mean excess function exists and is given by $e(v)=\frac{\beta+\xi(v-u)}{1-\xi}$, again, grows linearly in v

Sample mean excess plot and choice of the threshold

Definition (Sample Mean Excess Func; Mean Excess Plot)

For given loss data $X_1,\ldots,X_n$, the sample mean excess function is given by $e_n(v)=\frac{\sum_{i=1}^n(X_i-v)1_{\{X_i>v\}}}{\sum_{i=1}^n 1_{\{X_i>v\}}}$, and the mean excess plot is the plot of $\left(X_{(i)},e_n\left(X_{(i)}\right)\right), i=1,\ldots,n$, where $X_{(1)}\leq\ldots\leq X_{(n)}$ are the ordered loss data

- If the data supports the GPD model over u, $e_n(v)$ should become increasingly linear for higher values of $v\geq u$
- An upward, zero or downward trend indicates whether $\xi>0,\xi=0$ or $\xi<0$
- Select u as the smallest point where $e_n(v), v\geq u$, becomes linear
- Rule-of-thumb: One needs a couple of thousand data points and can often take u around the 90% quantile
- The sample mean excess plot is rarely perfectly linear (particularly for large u where one averages over a small number of excesses).

Modeling tails and measures of tail risk

- Let $N_u=\sum_{i=1}^n 1_{\{X_i>u\}}$ be the random number of exceedances of u by an iid sample $X_1,\ldots,X_n$ For $x\geq u$, one has

$$\bar{F}(x)=\mathbb{P}[X_1>x]=\mathbb{P}[X_1>u]\times\mathbb{P}[X_1>x\mid X_1>u]$$
$$=\mathbb{P}[X_1>u]\times\mathbb{P}[X_1-u>x-u\mid X_1>u]=\bar{F}(u)\bar{F}_u(x-u)$$

- Estimate $\bar{F}(u)$ empirically by N_u/n and $\bar{F}_u(x-u)$ by $1-G_{\hat{\xi},\hat{\beta}}(x-u)$
- The (semi-parametric) tail estimator of Smith

$$\hat{\bar{F}}(x)=\begin{cases}(N_u/n)\left(1+\hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}} & \hat{\xi}\neq 0,\\ (N_u/n)\exp(-(x-u)/\hat{\beta}) & \hat{\xi}=0,\end{cases}\quad x>u$$

- Bias-variance tradeoff: A high u reduces bias in estimating the excess function but increases the variance in estimating $F(u)$

GPD based VaR and ES estimates

For $\alpha>1-N_u/n$ and $\hat{\xi}\neq 0$, $\widehat{\mathrm{VaR}}_\alpha(X)=u+\frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{1-\alpha}{N_u/n}\right)^{-\hat{\xi}}-1\right)$

For $\alpha>1-N_u/n$ and $\hat{\xi}\in(-\infty,0)\cup(0,1)$, $\widehat{\mathrm{ES}}_\alpha(X)=e\left(\widehat{\mathrm{VaR}}_\alpha(X)\right)+\widehat{\mathrm{VaR}}_\alpha(X)=\frac{\hat{\beta}+\hat{\xi}\left(\widehat{\mathrm{VaR}}_\alpha(X)-u\right)}{1-\hat{\xi}}+\widehat{\mathrm{VaR}}_\alpha(X)=\frac{\widehat{\mathrm{VaR}}_\alpha(X)}{1-\hat{\xi}}+\frac{\hat{\beta}-\hat{\xi}u}{1-\hat{\xi}}$

- $\widehat{\mathrm{VaR}}_\alpha(X)$ usually outperforms the empirical quantile estimator $X_{([\alpha n])}$ when estimating at the edge of the sample
- Confidence intervals for $F(x)$, VaR, ES can be derived based on likelihood ratio test

6 Multivariate Models

6.1 Basics of Multivariate Modeling

Random Vectors and Their Distributions

Let $\mathbf{X}=(X_1,\ldots,X_d):\Omega\to\mathbb{R}^d$ be a d-dim random vector. The distribution of X is completely specified by the cdf $F_{\mathbf{X}}:\mathbb{R}^d\to[0,1]$, $F_{\mathbf{X}}(\mathbf{x})=\mathbb{P}[\mathbf{X}\leq\mathbf{x}]=\mathbb{P}[X_1\leq x_1,\ldots,X_d\leq x_d],\mathbf{x}\in\mathbb{R}^d$. It satisfies

(1) Normalization: $\lim_{x_1,\ldots,x_d\to\infty}F_X(x_1,\ldots,x_d)=1$ and $\lim_{x_j\to-\infty}F_X\left(x_1,\ldots,x_j,\ldots,x_d\right)=0$ for every $j=1,\ldots,d$

(2) Right-continuity: $F_{\mathbf{X}}(\mathbf{x}^n)\downarrow F_{\mathbf{X}}(\mathbf{x})$ for $\mathbf{x}^n\downarrow\mathbf{x}\in\mathbb{R}^d$

(3) d-Monotonicity: For all $\mathbf{a}\leq\mathbf{b}\in\mathbb{R}^d$,

$$\Delta_{(\mathbf{a},\mathbf{b}]}F_{\mathbf{X}}=\sum_{i\in\{0,1\}^d}(-1)^{\sum_{j=1}^d i_j}F_{\mathbf{X}}\left(a_1^{i_1}b_1^{1-i_1},\ldots,a_d^{i_d}b_d^{1-i_d}\right)\geq 0$$

On the other hand, every func $F:\mathbb{R}^d\to[0,1]$ satisfying (1)-(3) is a cdf of a d-dimensional random vec $\mathbf{X}$ (Carathéodory's extension theorem)

- The j-th marginal cdf of $\mathbf{X}$ is the cdf of X_j :

$$F_j\left(x_j\right)$$
$$=\mathbb{P}\left[X_j\leq x_j\right]=\mathbb{P}\left[X_1\leq\infty,\ldots,X_j\leq x_j,\ldots,X_d\leq\infty\right]$$
$$=F_{\mathbf{X}}\left(\infty,\ldots,x_j,\ldots,\infty\right),x_j\in\mathbb{R}$$

i.e., taking limits on all other elements.

- Similarly, the k-dimensional marginal cdf of X corresponding to $1\leq j_1\leq j_2\leq\ldots\leq j_k\leq d$ is the cdf of $\mathbf{Y}=\left(X_{j_1},\ldots,X_{j_k}\right)$:

$$F_{\mathbf{Y}}(\mathbf{y})=F_{\mathbf{X}}(\infty,y_1,\infty,\ldots,\infty,y_k,\infty),\mathbf{y}\in\mathbb{R}^k$$

- $F_{\mathbf{X}}$ is absolutely continuous if

$$F_{\mathbf{X}}(\mathbf{x})=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_d}f_{\mathbf{X}}(z_1,\ldots,z_d)dz_d\ldots dz_1$$
$$=\int_{(-\infty,\mathbf{x}]}f_{\mathbf{X}}(\mathbf{z})d\mathbf{z}$$

for some measurable function $f_{\mathbf{X}}:\mathbb{R}^d\to\mathbb{R}_+$, known as the pdf, or density.

- If $\mathbf{X}$ has a density $f_{\mathbf{X}}$, every marginal X_j has a density f_j, given by

$$f_j\left(x_j\right)=\int_{\mathbb{R}^{d-1}}f_{\mathbf{X}}\left(z_1,\ldots,x_j,\ldots,z_d\right)dz_d\ldots dz_1$$

- Existence of a density $\Rightarrow$ Existence of marginal densities for all k-dimensional marginals, $1\leq k\leq d-1$. The converse is false in general!
- The survival function $\bar{F}_{\mathbf{X}}$ of $\mathbf{X}$:

$$\bar{F}_{\mathbf{X}}(\mathbf{x})=\mathbb{P}[\mathbf{X}>\mathbf{x}]=\mathbb{P}[X_1>x_1,\ldots,X_d>x_d],\mathbf{x}\in\mathbb{R}^d$$

with corresponding j-th marginal survival func

$$\bar{F}_j\left(x_j\right)=\mathbb{P}\left[X_j>x_j\right]=\bar{F}_{\mathbf{X}}\left(-\infty,\ldots,x_j,\ldots,-\infty\right)$$

- Note that $\bar{F}_{\mathbf{X}}(\mathbf{x})\neq 1-F_X(x)$ (unless $d=1$)

Conditional distributions and independence

Denote $\mathbf{Y}_1 = (X_1,\ldots,X_k)$ and $\mathbf{Y}_2 = (X_{k+1},\ldots,X_d)$. The **conditional cdf** of $\mathbf{Y}_2$ given $\mathbf{Y}_1 = \mathbf{y}_1$ is

$$F_{\mathbf{Y}_2|\mathbf{Y}_1}(\mathbf{y}_2 \mid \mathbf{y}_1) = \mathbb{P}_{\mathbf{y}_1}[(-\infty,\mathbf{y}_2]],$$

where $\mathbb{P}_{\mathbf{y}_1}$ is a **regular conditional dist** of $\mathbf{Y}_2$ given $\mathbf{Y}_1$; that is:

(1) For every $y_1 \in \mathbb{R}^k, \mathbb{P}_{y_1}$ is a probability measure on $\mathbb{R}^{d-k}$

(2) For every Borel set B in $\mathbb{R}^{d-k}, \mathbb{P}_{y_1}[B]$ is measurable in y_1

(3) $\mathbb{P}_{\mathbf{Y}_1}[(-\infty,\mathbf{y}_2]] = \mathbb{E}\left[1_{\{\mathbf{Y}_2 \le \mathbf{y}_2\}} \mid \mathbf{Y}_1\right]$ a.s.

- **Disintegrate**: $F_{\mathbf{X}}(\mathbf{y}_1,\mathbf{y}_2) = \int_{(-\infty,\mathbf{y}_1]} F_{\mathbf{Y}_2|\mathbf{Y}_1}(\mathbf{y}_2 \mid z)\, dF_{\mathbf{Y}_1}(\mathbf{z})$
- For $\mathbf{y}_1 \to \infty$, one obtains $F_{\mathbf{Y}_2}(\mathbf{y}_2) = \int_{\mathbb{R}^k} F_{\mathbf{Y}_2|\mathbf{Y}_1}(\mathbf{y}_2 \mid \mathbf{z})\, dF_{\mathbf{Y}_1}(\mathbf{z})$
- If $\mathbf{X}$ has a density $f_{\mathbf{X}}$, then $f_{\mathbf{X}}(\mathbf{y}_1,\mathbf{y}_2) = \frac{\partial^2}{\partial \mathbf{y}_1 \partial \mathbf{y}_2} F_{\mathbf{X}}(\mathbf{y}_1,\mathbf{y}_2)$ and $f_{\mathbf{Y}_2}(\mathbf{y}_2) = \int_{\mathbb{R}^k} f_{\mathbf{X}}(\mathbf{z},\mathbf{y}_2)\, d\mathbf{z} = \int_{\mathbb{R}^k} f_{\mathbf{Y}_2|\mathbf{Y}_1}(\mathbf{y}_2 \mid \mathbf{z}) f_{\mathbf{Y}_1}(\mathbf{z}) d\mathbf{z}$, where $f_{\mathbf{Y}_2|\mathbf{Y}_1}(\mathbf{y}_2 \mid \mathbf{y}_1) = \frac{f_{\mathbf{X}}(\mathbf{y}_1,\mathbf{y}_2)}{f_{\mathbf{Y}_1}(\mathbf{y}_1)}$ is the **conditional density** of $\mathbf{Y}_2$ given $\mathbf{Y}_1 = \mathbf{y}_1$
- If $\mathbf{X}$ has a density $f_{\mathbf{X}}$, the **conditional cdf** can be recovered from the **conditional density**: $F_{\mathbf{Y}_2|\mathbf{Y}_1}(\mathbf{y}_2 \mid \mathbf{y}_1) = \int_{(-\infty,\mathbf{y}_2]} f_{\mathbf{Y}_2|\mathbf{Y}_1}(\mathbf{z} \mid \mathbf{y}_1)\, d\mathbf{z}$
- $\mathbf{Y}_1,\mathbf{Y}_2$ are **independent** $\Leftrightarrow$ $F_{\mathbf{X}}(\mathbf{y}_1,\mathbf{y}_2) = F_{\mathbf{Y}_1}(\mathbf{y}_1) F_{\mathbf{Y}_2}(\mathbf{y}_2)$ for all $\mathbf{y}_1 \in \mathbb{R}^k, \mathbf{y}_2 \in \mathbb{R}^{d-k}$
- If $\mathbf{X}$ has a density, then $\mathbf{Y}_1,\mathbf{Y}_2$ are **independent** $\Leftrightarrow f_{\mathbf{X}}(\mathbf{y}_1,\mathbf{y}_2) = f_{\mathbf{Y}_1}(\mathbf{y}_1) f_{\mathbf{Y}_2}(\mathbf{y}_2)$ for all $\mathbf{y}_1,\mathbf{y}_2$
- The components $X_1,\ldots,X_d$ of X are **independent** $\Leftrightarrow F_{\mathbf{X}}(\mathbf{x}) = \prod_{j=1}^d F_j(x_j)$ for all $\mathbf{x} \in \mathbb{R}^d$
- If $\mathbf{X}$ has a density, the components $X_1,\ldots,X_d$ are **independent** $\Leftrightarrow f_{\mathbf{X}}(\mathbf{x}) = \prod_{j=1}^d f_j(x_j)$ for all $\mathbf{x} \in \mathbb{R}^d$

Moments

- If $\mathbb{E}|X_j| < \infty$, for all $j = 1,\ldots,d$, the **mean vector** is $\mathbb{E}\mathbf{X} = (\mathbb{E}X_1,\ldots,\mathbb{E}X_d)^\top$ (column vec). If $X_1,\ldots,X_d$ are independent, then $\mathbb{E}[X_1 \cdots X_d] = \prod_{j=1}^d \mathbb{E}X_j$
- If $\mathbb{E}\left[X_j^2\right] < \infty$ for all $j = 1,\ldots,d$, the **covariance matrix** is $\Sigma = \mathrm{Cov}(\mathbf{X}) = \mathbb{E}\left[(\mathbf{X}-\mathbb{E}\mathbf{X})(\mathbf{X}-\mathbb{E}\mathbf{X})^\top\right]$. $\Sigma_{ij} = \mathrm{Cov}(\mathbf{X}) = \mathrm{Cov}(X_i,X_j), \Sigma_{jj} = \mathrm{Var}(X_j)$
- X_1, X_2 **ind** $\Rightarrow \mathrm{Cov}(X_1,X_2) = 0$. **The converse is not true!**
- If $\mathbb{E}\left[X_j^2\right] < \infty$ for all $j = 1,\ldots,d$, the **correlation matrix** $\mathrm{corr}(\mathbf{X})$ is $\mathrm{corr}(X_i,X_j) = \frac{\mathrm{Cov}(X_i,X_j)}{\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)}}$ (if $\mathrm{Var}(X_i)\mathrm{Var}(X_j) > 0$)
- By **Cauchy-Schwarz inequality**, $-1 \le \mathrm{corr}(X_i,X_j) \le 1$ with $\mathrm{corr}(X_i,X_j) = \pm 1$ iff $X_i = aX_j + b$ a.s. for $a \neq 0$ and $b \in \mathbb{R}$
- $\mathrm{Cov}(\mathbf{AX}+\mathbf{b}) = \mathbf{A}\Sigma\mathbf{A}^\top$
- Σ is PD (that is, $\mathbf{v}^\top \Sigma \mathbf{v} > 0$ for all $\mathbf{v} \in \mathbb{R}^d \backslash \{0\}$) $\Leftrightarrow$ all eigenvalues of Σ are positive $\Leftrightarrow \Sigma$ is invertible

Cholesky Decomposition

- A symmetric PD(PSD) Σ can be written as $\Sigma = \mathbf{A}\mathbf{A}^\top$ (**Cholesky decomp**), for a lower triangular $d \times d$-matrix $\mathbf{A}$ with $A_{jj} > 0\ (A_{jj} \ge 0)$
- If Σ is PD, the Cholesky decomp is **unique**. Otherwise, it is not.
- Consider a d-dimensional random vec $\mathbf{X}$ with iid standard normal components $X_1,\ldots,X_d$. Then $\mathrm{Var}(X_j) = 1$, and $\mathrm{Cov}(X_i,X_j) = 0$ for $i \neq j$.

  As a consequence, $\mathrm{Cov}(\mathbf{AX}) = \mathbf{A}\,\mathrm{Cov}(\mathbf{X})\mathbf{A}^\top = \mathbf{A}\mathbf{A}^\top = \Sigma \rightsquigarrow$ **every PSD matrix Σ is a cov matrix**

Characteristic Functions

The **characteristic function (cf)** of a d-dimensional random vec $\mathbf{X}$ is the func $\phi_{\mathbf{X}} : \mathbb{R}^d \to \mathbb{C}$, $\phi_{\mathbf{X}}(\mathbf{u}) = \mathbb{E}\left[\exp\left(i\mathbf{u}^T\mathbf{X}\right)\right]$ (for a complex-valued RV $Z = V + iW$, one defines $\mathbb{E}Z = \mathbb{E}V + i\mathbb{E}W$)
The cf is **determined by and determines** the distribution of a random vector.
The components $X_1,\ldots,X_d$ of a random vec $\mathbf{X}$ are independent $\Leftrightarrow \phi_{\mathbf{X}}(\mathbf{u}) = \prod_{j=1}^d \phi_{X_j}(u_j)$ for all $\mathbf{u} \in \mathbb{R}^d$

Standard Estimators of Means, Cov & Corr

Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ be **uncorrelated** d-dim random vecs, all with the same cdf F Assume second moments exist and set $\mu = \mathbb{E}\mathbf{X}_1, \Sigma = \mathrm{Cov}(\mathbf{X}_1), \mathbf{P} = \mathrm{corr}(\mathbf{X}_1)$.

- **Sample means**: $\overline{X}_j = \frac{1}{n}\sum_{t=1}^n X_{t,j}$, **unbiased**
- **Sample cov**: $S_{ij} = \frac{1}{n}\sum_{t=1}^n \left(X_{t,i} - \overline{X}_i\right)\left(X_{t,j} - \overline{X}_j\right)$
- **Sample corr**: $R_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}}$
- S is **biased**. But $S^n = \frac{n}{n-1}S$ is **unbiased**.

  Proof: Since there is no serial correlation,

$$\mathbb{E}S_{ij}^n = \frac{1}{n-1}\mathbb{E}\sum_{t=1}^n \left(X_{t,i} - \overline{X}_i\right)\left(X_{t,j} - \overline{X}_j\right)$$

$$= \frac{1}{n-1}\mathbb{E}\sum_{t=1}^n \left(\left[X_{t,i} - \mu_i\right] - \left[\overline{X}_i - \mu_i\right]\right)\left(\left[X_{t,j} - \mu_j\right] - \left[\overline{X}_j - \mu_j\right]\right)$$

$$= \frac{1}{n-1}\left(n\Sigma_{ij} - \frac{2n}{n}\Sigma_{ij} + \frac{n^2}{n^2}\Sigma_{ij}\right) = \Sigma_{ij}$$

Normal Distributions

- A random vec $\mathbf{Z}$ with components $Z_1,\ldots,Z_d$ is **d-dim standard normal** if $Z_1,\ldots,Z_d$ are ind one-dimensional standard normal, or equivalently, if it has density $\prod_{j=1}^d \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x_j^2}{2}\right) = \frac{1}{(2\pi)^{d/2}}\exp\left(-\frac{1}{2}\sum_{j=1}^d x_j^2\right)$
- A d-dim random vec $\mathbf{X}$ has a **multivariate normal (or Gaussian) distribution** if $\mathbf{X} = \boldsymbol{\mu} + \mathbf{AZ}$, where $\boldsymbol{\mu} \in \mathbb{R}^d, \mathbf{A} \in \mathbb{R}^{d \times k}$ and $\mathbf{Z}$ is k-dim standard normal. $\mathbb{E}\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbb{E}\mathbf{Z} = \boldsymbol{\mu}$, $\mathrm{Cov}(\mathbf{X}) = \mathbf{A}\,\mathrm{Cov}(\mathbf{Z})\mathbf{A}^\top = \mathbf{A}\mathbf{A}^\top =: \Sigma$
- If X is one-dim standard normal, then $\mathbb{E}\exp(aX) = \exp\left(\frac{a^2}{2}\right) \rightsquigarrow \phi_X(u) = \mathbb{E}\exp(iuX) = \exp\left(-\frac{u^2}{2}\right)$
- If $\mathbf{X}$ is d-dim standard normal, then $\phi_{\mathbf{X}}(\mathbf{u}) = \exp\left(-\frac{1}{2}\mathbf{u}^\top\mathbf{u}\right)$
- If $\mathbf{X} = \boldsymbol{\mu} + \mathbf{AZ}$, then $\phi_{\mathbf{X}}(\mathbf{u}) = \mathbb{E}\exp\left(i\mathbf{u}^\top\mathbf{X}\right) = \exp\left(i\mathbf{u}^\top\boldsymbol{\mu}\right)\mathbb{E}\exp\left(i\mathbf{u}^\top\mathbf{AZ}\right) = \exp\left(i\mathbf{u}^\top\mu - \frac{1}{2}\mathbf{u}^\top\Sigma\mathbf{u}\right)$

If $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu},\Sigma)$, then

- $\mathbf{Y} = \mathbf{v} + \mathbf{MX}$ for $\mathbf{v} \in \mathbb{R}^m$ and $\mathbf{M} \in \mathbb{R}^{m \times d}$ is $\mathcal{N}_m\left(\mathbf{v} + \mathbf{M}\boldsymbol{\mu}, \mathbf{M}\Sigma\mathbf{M}^\top\right)$
- In particular, $\mathbf{Y} = \left(X_{j_1},\ldots,X_{j_m}\right)$ is $\mathcal{N}_m(\mathbb{E}\mathbf{Y}, \mathrm{Cov}(\mathbf{Y}))$
- **Margins**: $X_j \sim \mathcal{N}\left(\mu_j,\Sigma_{jj}\right)$; **Sum**: $\sum_{j=1}^d X_j \sim \mathcal{N}\left(\sum_{j=1}^d \mu_j, \sum_{i,j=1}^d \Sigma_{ij}\right)$
- X_i and X_j are ind $\Leftrightarrow \phi_{(X_i,X_j)}(x_i,x_j) = \phi_{X_i}(x_i)\phi_{X_j}(x_j) \Leftrightarrow \mathrm{Cov}(X_i,X_j) = 0$
- $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu},\Sigma) \Leftrightarrow \mathbf{v}^\top\mathbf{X} = \mathcal{N}\left(\mathbf{v}^\top\boldsymbol{\mu}, \mathbf{v}^\top\Sigma\mathbf{v}\right)$ for all $\mathbf{v} \in \mathbb{R}^d$
- We call $\mathbf{X}$ **regular normal** if Σ is invertible and **singular normal** if Σ is not invertible.
- $\mathbf{X}$ has a density **iff** it is regular, which is $\mathbf{x} \in \mathbb{R}^d$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det\Sigma}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

Proof: If Σ is invertible, then $\mathbf{X} \overset{(d)}{=} \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$ for an invertible $\mathbf{A} \in \mathbb{R}^{d \times d}$ such that $\mathbf{A}\mathbf{A}^\top = \Sigma$.

So for $\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} \leftrightarrow \mathbf{z} = \mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\mu})$:

$$\mathbb{P}[\mathbf{X} \in B] = \mathbb{P}[\boldsymbol{\mu} + \mathbf{A}\mathbf{Z} \in B] = \mathbb{P}\left[\mathbf{Z} \in \mathbf{A}^{-1}(B-\boldsymbol{\mu})\right]$$

$$= \frac{1}{(2\pi)^{\frac{d}{2}}}\int_{\mathbf{A}^{-1}(B-\mu)} \exp\left(-\frac{\mathbf{z}^\top\mathbf{z}}{2}\right) d\mathbf{z}$$

$$= \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det\Sigma}}\int_B \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^\top\left(\mathbf{A}^{-1}\right)^\top\mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right.$$

$$= \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det\Sigma}}\int_B \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right) d\mathbf{x}$$

- The **contour sets** of the above density consist of all $x \in \mathbb{R}^d$ satisfying $(\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \mathbf{c}$. **The contour sets are ellipsoids**.
- More generally, distributions with densities whose contour sets are ellipsoids are called **elliptical**.

Sampling from a $\mathcal{N}_d(\boldsymbol{\mu},\Sigma)$ distribution

(1) Compute Cholesky decomposition: $\Sigma = \mathbf{A}\mathbf{A}^\top$

(2) Generate ind standard normals $Z_1,\ldots,Z_d$

(3) Return $\mathbf{X} = \boldsymbol{\mu} + \mathbf{AZ}$

Conditioning normal distributions

Let $X \sim \mathcal{N}_d(\boldsymbol{\mu},\Sigma)$ with Σ PD. Denote $\mathbf{Y}_1 = (X_1,\ldots,X_k), \mathbf{Y}_2 = (X_{k+1},\ldots,X_d)$, $\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}$ and $\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$ (note: Σ_{11},Σ_{22} are symmetric and $\Sigma_{21} = \Sigma_{12}^\top$).
One has $(\mathbf{Y}_2 \mid \mathbf{Y}_1 = \mathbf{y}_1) \sim \mathcal{N}_{d-k}(\tilde{\boldsymbol{\mu}},\tilde{\Sigma})$, where $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$, and $\tilde{\Sigma} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$

Proof for $d = 2$ and $k = 1$

$\mathbf{X} = \begin{pmatrix}X_1\\ X_2\end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \Sigma = \begin{pmatrix}a & b\\ b & c\end{pmatrix}, \Sigma^{-1} = \frac{1}{ac-b^2}\begin{pmatrix}c & -b\\ -b & a\end{pmatrix}$. Then

$$f_{X_2|X_1}(x_2 \mid x_1) = \frac{f_X(x_1,x_2)}{f_{X_1}(x_1)}$$

$$= C\frac{\exp\left(-(x-\mu)^T\Sigma^{-1}(x-\mu)/2\right)}{\exp\left(-a^{-1}(x_1-\mu_1)^2/2\right)}$$

$$= C\exp\left\{\left(\frac{1}{a} - \frac{c}{ac-b^2}\right)\frac{(x_1-\mu_1)^2}{2}\right.$$

$$\left. + \frac{b}{ac-b^2}(x_1-\mu_1)(x_2-\mu_2) - \frac{a}{ac-b^2}\frac{(x_2-\mu_2)^2}{2}\right\}$$

$$= h(x_1)\exp\left(-\frac{a}{2\left(ac-b^2\right)}\left(x_2 - \left[\mu_2 + \frac{b}{a}(x_1-\mu_1)\right]\right)^2\right)$$

So $(X_2 \mid X_1 = x_1) \sim \mathcal{N}\left(\mu_2 + \frac{b}{a}(x_1-\mu_1), c - \frac{b^2}{a}\right.$

Convolutions

If $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ and $\mathbf{Y} \sim \mathcal{N}_d(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})$ are ind, then $\mathbf{X}+\mathbf{Y} = \mathcal{N}_d(\boldsymbol{\mu}+\tilde{\boldsymbol{\mu}}, \Sigma+\tilde{\Sigma})$

Quadratic forms

Let $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ with Σ PD and $\mathbf{A} \in \mathbb{R}^{d\times d}$ such that $\mathbf{A}\mathbf{A}^\top = \Sigma$ Then $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$ for $\mathbf{Z} \sim \mathcal{N}_d(0, \mathbf{I}_d)$.
So $\mathbf{Z} = \mathbf{A}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \mathcal{N}_d(0, \mathbf{I}_d)$ and $(\mathbf{X}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu}) = (\mathbf{X}-\boldsymbol{\mu})^\top \left(\mathbf{A}^{-1}\right)^\top \mathbf{A}^{-1}(\mathbf{X}-\boldsymbol{\mu}) = \mathbf{Z}^\top\mathbf{Z} \sim \chi^2_d$

Testing multivariate normality

If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are iid $\mathcal{N}_d(\mu, \Sigma)$, then, for $\mathbf{a} \in \mathbb{R}^d$, $\mathbf{a}^\top\mathbf{X}_1, \ldots, \mathbf{a}^\top\mathbf{X}_n$ are iid $\mathcal{N}\left(\mathbf{a}^\top\boldsymbol{\mu}, \mathbf{a}^\top\Sigma\mathbf{a}\right)$ This can be tested statistically (for different $\mathbf{a}$) with various goodness-of-fit tests (e.g. Q-Q plots)

Mardia's test

- If $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ with Σ PD, then $(\mathbf{X}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi^2_d$
- Let $D_i^2 = \left(\mathbf{X}_i - \bar{\mathbf{X}}\right)^\top \Sigma^{-1}\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)$ denote the **squared Mahalanobis distances** and $D_{ij} = \left(\mathbf{X}_i - \bar{\mathbf{X}}\right)^\top \Sigma^{-1}\left(\mathbf{X}_j - \bar{\mathbf{X}}\right)$ the **Mahalanobis angles**
- Set $b_d = \frac{1}{n^2}\sum_{i,j=1}^n D_{ij}^3$ and $k_d = \frac{1}{n}\sum_{i=1}^n D_i^4$
- Under the **null hypothesis**, one has $\frac{n}{6}b_d \to \chi^2_{d(d+1)(d+2)/6}$ and $\frac{k_d - d(d+2)}{\sqrt{8d(d+2)/n}} \to \mathcal{N}(0,1)$ for $n \to \infty$,

Advantages & Drawbacks of $N_d(\mu, \Sigma)$

Advantages: Inference easy; Distribution is determined by μ and Σ; Linear combinations are normal $\rightsquigarrow$ VaR and ES calculations for portfolios are easy; Marginal distributions are normal; Conditional distributions are normal; Quadratic forms are known; Convolutions are normal; Simulation is straightforward; Independence and uncorrelatedness are equivalent

Drawbacks:

(1) Tails of univariate (normal) margins are too thin (generate too few extreme events)

(2) Joint tails are too thin (generate too few joint extreme events). $\mathcal{N}_d(\mu, \Sigma)$ cannot capture the notion of tail dependence.

(3) Very strong symmetry: radial symmetry.

In short:

- Normal variance mixture distributions can address (1) and (2) while sharing many of the desirable properties of $\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$
- Normal mean-variance mixture distributions can also address (3) but at the expense of tractability in comparison to $\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$

6.2 Normal mixture distributions

Multivariate normal variance mixtures

A d-dim random vec $\mathbf{X}$ has a (multivariate) normal variance mixture distribution if $\mathbf{X} \overset{(d)}{=} \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\mathbf{Z}$, where $\boldsymbol{\mu} \in \mathbb{R}^d, \mathbf{A} \in \mathbb{R}^{d\times k}, \mathbf{Z} \sim \mathcal{N}_k(0, \mathbf{I}_k)$ and $W \geq 0$ is a RV ind of $\mathbf{Z}$. μ is called **location vec** and $\Sigma = \mathbf{A}\mathbf{A}^\top$ **scale (or dispersion) matrix**.

Note: $(\mathbf{X} \mid W = w) \overset{(d)}{=} \boldsymbol{\mu} + \sqrt{w}\mathbf{A}\mathbf{Z} \sim \mathcal{N}_d\left(\boldsymbol{\mu}, w\mathbf{A}\mathbf{A}^\top\right) = \mathcal{N}_d(\boldsymbol{\mu}, w\Sigma)$ or $\mathbf{X} \mid W \sim \mathcal{N}_d(\boldsymbol{\mu}, W\Sigma)$.
W **can be interpreted as a shock** affecting the variances of all risk factors

Properties of multivariate normal variance mixtures

Let $\mathbf{X} = \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\mathbf{Z}$ and $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$

- If $\mathbb{E}\sqrt{W} < \infty$, then $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu} + \mathbb{E}[\sqrt{W}]\mathbf{A}\mathbb{E}[\mathbf{Z}] = \boldsymbol{\mu} = \mathbb{E}[\mathbf{Y}]$
- If $\mathbb{E}W < \infty$, then $\operatorname{Cov}(\mathbf{X}) = \operatorname{Cov}(\sqrt{W}\mathbf{A}\mathbf{Z}) = \mathbb{E}\left[W\mathbf{A}\mathbf{Z}\mathbf{Z}^\top\mathbf{A}^\top\right] = \mathbb{E}[W]\mathbf{A}\,\mathbb{E}\left[\mathbf{Z}\mathbf{Z}^\top\right]\mathbf{A}^\top = \mathbb{E}[W]\Sigma \overset{\text{(in general)}}{\neq} \Sigma = \operatorname{Cov}(\mathbf{Y})$
- If $\mathbb{E}W < \infty$, then $\operatorname{corr}(\mathbf{X}) = \operatorname{corr}(\mathbf{Y})$

Lemma (Independence in normal variance mixtures)

Let $\mathbf{X} = \mu + \sqrt{W}\mathbf{Z}$ with $\mathbb{E}W < \infty$ (uncorrelated normal variance mixture). Then X_i and X_j are independent $\Leftrightarrow$ W is a.s. constant (i.e. $\mathbf{X} \sim \mathcal{N}_d$)

- **Recall**: cf of $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ is $\phi_{\mathbf{X}}(\mathbf{u}) = \exp\left(i\,\mathbf{u}^\top\boldsymbol{\mu} - \frac{1}{2}\mathbf{u}^\top\Sigma\mathbf{u}\right)$
- $(\mathbf{X} \mid W = w) \sim \mathcal{N}_d(\boldsymbol{\mu}, w\Sigma)$
- The **characteristic function** of a multivariate normal variance mixture is $\phi_{\mathbf{X}}(\mathbf{u}) = \mathbb{E}\left[\mathbb{E}\left[\exp\left(i\mathbf{u}^\top\boldsymbol{\mu} + i\mathbf{u}^\top\sqrt{W}\mathbf{A}\mathbf{Z}\right) \mid W\right]\right] = \exp\left(i\mathbf{u}^\top\boldsymbol{\mu}\right)\mathbb{E}\exp\left(-W\frac{1}{2}\mathbf{u}^\top\Sigma\mathbf{u}\right)$

Laplace–Stieltjes tranform

The Laplace-Stieltjes transform of F_W is $\hat{F}_W(\theta) = \mathbb{E}\left[e^{-\theta W}\right] = \int_0^\infty e^{-\theta w}\,dF_W(w)$.
Therefore $\phi_{\mathbf{X}}(u) = \exp\left(i\mathbf{u}^\top\boldsymbol{\mu}\right)\hat{F}_W\left(\frac{1}{2}\mathbf{u}^\top\Sigma\mathbf{u}\right)$

Notation: $\mathbf{X} \sim \mathcal{M}_d\left(\boldsymbol{\mu}, \Sigma, \hat{F}_W\right)$ for a d-dimensional multivariate normal variance mixture

Density

If Σ is PD and $\mathbb{P}[W = 0] = 0$, the density of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \int_0^\infty f_{\mathbf{X}|W}(x \mid w)\,dF_W(w)$$

$$= \int_0^\infty \frac{1}{(2\pi w)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2w}\right)dF_W(w)$$

- Only depends on $\mathbf{x}$ through $(\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})$
- If Σ is diagonal and $\mathbb{E}W < \infty$, the components of $\mathbf{X}$ are uncorrelated (as $\operatorname{Cov}(\mathbf{X}) = \mathbb{E}[W]\Sigma$) but not independent unless W is constant a.s.

Affine transformations

- For $\mathbf{X} \sim \mathcal{M}_d\left(\boldsymbol{\mu}, \Sigma, \hat{F}_W\right)$ and $\mathbf{Y} = \mathbf{b} + \mathbf{B}\mathbf{X}$, where $\mathbf{b} \in \mathbb{R}^k$ and $\mathbf{B} \in \mathbb{R}^{k\times d}$, one has $\mathbf{Y} \sim \mathcal{M}_k\left(\mathbf{b}+\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\Sigma\mathbf{B}^\top, \hat{F}_W\right)$

  Indeed, if $\mathbf{X} = \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\mathbf{Z}$, then $\mathbf{b}+\mathbf{B}\mathbf{X} = \mathbf{b}+\mathbf{B}\boldsymbol{\mu}+\sqrt{W}\mathbf{B}\mathbf{A}\mathbf{Z}$
- Particularly, $\mathbf{v}^\top\mathbf{X} \sim \mathcal{M}_1\left(\mathbf{v}^\top\boldsymbol{\mu}, \mathbf{v}^\top\Sigma\mathbf{v}, \hat{F}_W\right), \mathbf{v} \in \mathbb{R}^d$

Sampling / Simulation Algorithm of $\mathbf{X} = \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\mathbf{Z} \sim \mathcal{M}_d\left(\boldsymbol{\mu}, \Sigma, \hat{F}_W\right)$

(1) Generate $\mathbf{Z} \sim \mathcal{N}_d(0, \mathbf{I}_d)$

(2) Generate $W \sim F_W$, independent of $\mathbf{Z}$. e.g. $W = F_W^{-1}(U)$ for $U \sim \text{Unif}(0,1)$

(3) Compute a Cholesky decomposition $\Sigma = \mathbf{A}\mathbf{A}^\top$

(4) Return $\mathbf{X} = \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\mathbf{Z}$

Examples of multivariate normal mixtures

- **Multivariate normal**: $W \equiv 1$
- **Two point mixture**: $W = \begin{cases} w_1 & \text{with prob } p \\ w_2 & \text{with prob } 1-p \end{cases}$ can be used to model **an ordinary and a stress regime**
- **k point mixture**:

$$W = \begin{cases} w_1 & \text{with prob } p_1 \\ \vdots & \vdots \\ w_k & \text{with prob } p_k \end{cases}$$

Multivariate t-distribution

Multivariate t-distribution is also an example of multivariate normal mixtures.
Set $W = \nu/V$ for $V \sim \chi^2_\nu$, **or equivalently**, $W = 1/G$ for $G \sim \Gamma(\nu/2, \nu/2)$.
W has an **inverse Gamma distribution**.

- **Density** of the multivariate t-distribution is

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

  where $\boldsymbol{\mu} \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d\times d}$ is PD and $\nu > 0$ is the degrees of freedom.

  Notation: $\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$
- $t_d(\nu, \boldsymbol{\mu}, \Sigma)$ has heavier marginal and joint tails than $\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$
- If $\nu > 2$, then $\mathbb{E}[W] = \frac{\nu}{(\nu-2)}$, and so, $\operatorname{Cov}(\mathbf{X}) = \frac{\nu}{(\nu-2)}\Sigma$. For finite variances/correlations, $\nu > 2$ is required. For finite mean, $\nu > 1$ is required.

Multivariate normal mean-variance mixtures

A d-dim random vec $\mathbf{X}$ has a **(multivariate) normal mean-variance mixture distribution** if $\mathbf{X} \overset{(d)}{=} m(W) + \sqrt{W}\mathbf{A}\mathbf{Z}$, where $\mathbf{Z} \sim \mathcal{N}_k(0, \mathbf{I}_k)$; $\mathbf{A} \in \mathbb{R}^{d\times k}$; $W \geq 0$ is a RV ind of $\mathbf{Z}$; $m : \mathbb{R}_+ \to \mathbb{R}^d$ is a measurable function.

- Normal mean-var mixtures can add **skewness**
- Denote $\Sigma = \mathbf{A}\mathbf{A}^\top$ and observe that $(\mathbf{X} \mid W = w) \sim \mathcal{N}_d(m(w), w\Sigma)$
- In general, they are **no longer elliptical**

Example: Suppose $m(W) = \boldsymbol{\mu} + \boldsymbol{\gamma}W$ for $\mu, \gamma \in \mathbb{R}^d$.
Since $\mathbb{E}[\mathbf{X} \mid W] = \boldsymbol{\mu} + \boldsymbol{\gamma}W$ and $\operatorname{Cov}(\mathbf{X} \mid W) = W\Sigma$, one has $\mathbb{E}\mathbf{X} = \mathbb{E}[\mathbb{E}[\mathbf{X} \mid W]] = \boldsymbol{\mu} + \boldsymbol{\gamma}\mathbb{E}[W]$ if $\mathbb{E}W < \infty$; $\operatorname{Cov}(\mathbf{X}) = \operatorname{Cov}(\mathbb{E}[\mathbf{X} \mid W]) + \mathbb{E}[\operatorname{Cov}(\mathbf{X} \mid W)] = \operatorname{Var}(W)\boldsymbol{\gamma}\boldsymbol{\gamma}^\top + \mathbb{E}[W]\Sigma$ if $\mathbb{E}W^2 < \infty$.
If W has a **GIG (generalized inverse Gaussian)** distribution, that is, it has density

$$f(x) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})}x^{(p-1)}e^{-(ax+b/x)/2}, \quad x > 0$$

for parameters $a, b > 0$ and $p \in \mathbb{R}$.
Then X has a **generalized hyperbolic distribution**.
K_p is a modified Bessel function of the second kind.
In the special case $\boldsymbol{\gamma} = \mathbf{0}$, one obtains an (elliptical) normal variance mixture.

6.3 Spherical and elliptical distributions

- $M_d\left(\mu, \Sigma, \hat{F}_W\right)$ (e.g. multivariate t, generalized hyperbolic) are better models than $\mathcal{N}_d(\mu, \Sigma)$ for **daily/weekly log-returns of stocks**.
- The more general **skewed normal mean-variance mixture distributions** offer only a modest improvement.
- Elliptical distributions are a generalization of $M_d\left(\mu, \Sigma, \hat{F}_W\right)$.

Spherical distributions

A d-dim random vec $\mathbf{Y}$ has a **spherical distribution** if for every orthogonal $\mathbf{U} \in \mathbb{R}^{d\times d}$ (i.e. $\mathbf{U}^\top\mathbf{U} = \mathbf{U}\mathbf{U}^\top = \mathbf{I}_d$), one has $\mathbf{Y} \overset{(d)}{=} \mathbf{U}\mathbf{Y}$ (**distributional invariance under rotations and reflections**).

Thm: Characterization of spherical distributions

Denote $\|\mathbf{u}\| = \sqrt{u_1^2 + \cdots + u_d^2}$. Then the following are **equivalent**: (1) $\mathbf{Y}$ is spherical; (2) For every $\mathbf{a} \in \mathbb{R}^d, \mathbf{a}^\top\mathbf{Y} \overset{(d)}{=} \|\mathbf{a}\|Y_1$; (3) There exists a **characteristic generator** $\psi : \mathbb{R}_+ \to \mathbb{R}$ such that $\phi_{\mathbf{Y}}(\mathbf{u}) = \psi\left(\|\mathbf{u}\|^2\right), \mathbf{u} \in \mathbb{R}^d$

Notation: $\mathbf{Y} \sim \mathcal{S}_d(\psi)$

Additivity

Let $\mathbf{Y}_i \sim \mathcal{S}_d(\psi_i), i = 1, \ldots, n$, be ind spherically distributed random vecs and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. If $\mathbf{Z} = \sum_{i=1}^n \alpha_i\mathbf{Y}_i$, then $\mathbf{Z} \sim \mathcal{S}_d(\psi)$ for $\psi(x) = \prod_{i=1}^n \psi_i\left(\alpha_i^2 x\right)$

Thm: Stochastic representation

$\mathbf{Y} \sim \mathcal{S}_d(\psi) \Leftrightarrow \mathbf{Y} \overset{(d)}{=} R\mathbf{S}$ for $\mathbf{S} \sim U\left(\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\right)$ and an independent radial part $R \geq 0$.
If $\mathbf{Y}$ has a density $f_\mathbf{Y}$, it must be of the form $f_\mathbf{Y}(\mathbf{y}) = g\left(\|\mathbf{y}\|^2\right)$ for a function $g : \mathbb{R}_+ \to \mathbb{R}_+$, referred to as density generator.

Corollary

If $\mathbf{Y} \sim \mathcal{S}_d(\psi)$ and $\mathbb{P}[\mathbf{Y} = 0] = 0$, then $\left(\|\mathbf{Y}\|, \frac{\mathbf{Y}}{\|\mathbf{Y}\|}\right) \overset{(d)}{=} \left(\|R\mathbf{S}\|, \frac{R\mathbf{S}}{\|R\mathbf{S}\|}\right) \overset{(d)}{=} (R, \mathbf{S})$.
In particular, $\|\mathbf{Y}\|$ and $\mathbf{Y}/\|\mathbf{Y}\|$ are independent.

Examples

- $\mathbf{Y} \sim \mathcal{M}_d\left(0, \mathbf{I}_d, \hat{F}_W\right)$ is spherical.

  Indeed, since $\mathbf{Y} \overset{(d)}{=} \sqrt{W}\mathbf{Z}$, one has $\phi_\mathbf{Y}(\mathbf{u}) = \mathbb{E}\left[\mathbb{E}\left[\exp\left(i\sqrt{W}\mathbf{u}^\top \mathbf{Z}\right) \mid W\right]\right] = \mathbb{E}\left[\exp\left(-W\|\mathbf{u}\|^2/2\right)\right] = \hat{F}_W\left(\|\mathbf{u}\|^2/2\right)$

  So $\mathbf{Y} \sim \mathcal{S}_d(\psi)$ for $\psi(x) = \hat{F}_W(x/2)$
- For $\mathbf{Y} \sim \mathcal{N}_d(0, \mathbf{I}_d), \psi(x) = \exp(-x/2)$. By the Corollary above, simulating $\mathbf{S} \sim U\left(\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\right)$ can thus be done through $\mathbf{S} = \mathbf{Y}/\|\mathbf{Y}\|$.

  Also, $R^2 = \mathbf{Y}^\top\mathbf{Y} \sim \chi^2_d$. So (1) $\mathbb{E}[R]\mathbb{E}[\mathbf{S}] = \mathbb{E}\mathbf{Y} = \mathbf{0} \rightsquigarrow \mathbb{E}\mathbf{S} = \mathbf{0}$ and (2) $d\,\mathrm{Cov}(\mathbf{S}) = \mathbb{E}\left[R^2\right]\mathrm{Cov}(\mathbf{S}) = \mathrm{Cov}(\mathbf{Y}) = \mathbf{I}_d \rightsquigarrow \mathrm{Cov}(\mathbf{S}) = \mathbf{I}_d/d$
- For $\mathbf{Y} \sim \mathcal{S}_d(\psi)$ with $\mathbb{E}\left[R^2\right] < \infty$, one has $\mathrm{Cov}(\mathbf{Y}) = \mathbb{E}\left[R^2\right]\mathrm{Cov}(\mathbf{S}) = \frac{\mathbb{E}[R^2]}{d}\mathbf{I}_d$, and $\mathrm{corr}(\mathbf{Y}) = \mathbf{I}_d$
- For $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$ with $\mathbb{E}\left[R^2\right] < \infty$ and Cholesky factor $\mathbf{A}$ of a covariance matrix Σ, one has $\mathrm{Cov}(\mathbf{X}) = \mathbb{E}\left[\mathbf{A}\mathbf{Y}\mathbf{Y}^\top\mathbf{A}^\top\right] = \mathbf{A}\,\mathrm{Cov}(\mathbf{Y})\mathbf{A}^\top = \frac{\mathbb{E}[R^2]}{d}\Sigma$ and $\mathrm{corr}(\mathbf{X}) = \mathbf{P}$ (corr matrix corresponding to Σ)

Example: t-distribution

For $\mathbf{Y} \sim t_d(\nu, 0, \mathbf{I}_d)$, one has $R^2 = \mathbf{Y}^\top\mathbf{Y} = W\mathbf{Z}^\top\mathbf{Z}$ for $W = \nu/V, V \sim \chi^2_\nu, \mathbf{Z} \sim \mathcal{N}_d(0, \mathbf{I}_d)$. So $\frac{R^2}{d} = \frac{\mathbf{Z}^\top\mathbf{Z}/d}{(\nu/W)/\nu} = \frac{\chi^2_d/d}{\chi^2_\nu/\nu} \sim F(d, \nu)$ (F-distribution). Therefore, $\mathbb{E}[W] = \mathbb{E}\left[R^2\right]/\mathbb{E}\left[\mathbf{Z}^\top\mathbf{Z}\right] = \mathbb{E}\left[R^2\right]/d = \nu/(\nu-2)$ (if $\nu > 2$).
It follows that $\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$ has $\mathrm{Cov}(\mathbf{X}) = \frac{\nu}{\nu-2}\Sigma$ and $\mathrm{corr}(\mathbf{X}) = \mathbf{P}$.
One can use a Q-Q-plot of the order statistics of $R^2/d = \|\mathbf{Y}\|^2/d$ versus the theoretical quantiles of a (hypothesized) $F(d, \nu)$-distribution to check the goodness-of-fit of a hypothesized t-distribution.

Elliptical distributions

A d-dim random vec $\mathbf{X}$ has an elliptical distribution if $\mathbf{X} \overset{(d)}{=} \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$ for $\boldsymbol{\mu} \in \mathbb{R}^d, \mathbf{A} \in \mathbb{R}^{d\times k}$ and $\mathbf{Y} \in \mathcal{S}_k(\psi)$
location vector: $\boldsymbol{\mu}$, scale matrix: $\Sigma = \mathbf{A}\mathbf{A}^\top$

- Elliptical random vec has stochastic representation $\mathbf{X} \overset{(d)}{=} \boldsymbol{\mu} + R\mathbf{A}\mathbf{S}$, where $R = \|\mathbf{Y}\|$ and $\mathbf{S} = \mathbf{Y}/\|\mathbf{Y}\|$
- The characteristic func of an elliptical random vec is $\phi_\mathbf{X}(\mathbf{u}) = \mathbb{E}\exp\left(i\mathbf{u}^\top\mathbf{X}\right) = e^{i\mathbf{u}^\top\boldsymbol{\mu}}\mathbb{E}\exp\left(i\mathbf{u}^\top\mathbf{A}\mathbf{Y}\right) = e^{i\mathbf{u}^T}\boldsymbol{\mu}\mathbb{E}\exp\left(i\left(\mathbf{A}^\top\mathbf{u}\right)^\top\mathbf{Y}\right) = e^{i\mathbf{u}^\top\boldsymbol{\mu}}\psi\left(\mathbf{u}^\top\mathbf{A}\mathbf{A}^\top\mathbf{u}\right) = e^{i\mathbf{u}^\top\boldsymbol{\mu}}\psi\left(\mathbf{u}^\top\Sigma\mathbf{u}\right)$

  Notation: $E_d(\boldsymbol{\mu}, \Sigma, \psi)$. Note that $E_d(\boldsymbol{\mu}, \Sigma, \psi) = E_d(\boldsymbol{\mu}, c\Sigma, \psi(\cdot/c))$
- If Σ is PD with Cholesky factor $\mathbf{A}$, then $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ iff $\mathbf{Y} = \mathbf{A}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim S_d(\psi)$, in which case, $\left(\sqrt{(\mathbf{X}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu})}, \frac{\mathbf{A}^{-1}(\mathbf{X}-\boldsymbol{\mu})}{\sqrt{(\mathbf{X}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu})}}\right) \overset{(d)}{=} (R, \mathbf{S})$, which can be used for testing elliptical symmetry.
- Normal variance mixture distributions are all elliptical since $\mathbf{X} \overset{(d)}{=} \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\mathbf{Z} = \boldsymbol{\mu} + \sqrt{W}\|\mathbf{Z}\|\mathbf{A}\frac{\mathbf{Z}}{\|\mathbf{Z}\|} = \boldsymbol{\mu} + R\mathbf{A}\mathbf{S}$ for $R = \sqrt{W}\|\mathbf{Z}\|$ and $\mathbf{S} = \mathbf{Z}/\|\mathbf{Z}\|$, where R and $\mathbf{S}$ are ind.

Properties of elliptical distributions

- Density: Let $\mathbf{A}\mathbf{A}^\top = \Sigma$ be PD and $\mathbf{Y} \sim \mathcal{S}_d(\psi)$ have density generator g. Then by the Density Transformation Theorem, $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$ has density

  $$f_\mathbf{X}(\mathbf{x}) = \frac{1}{\sqrt{\det\Sigma}}g\left((\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

  It depends on $\mathbf{x}$ only through $(\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})$.
  In particular, the level sets are ellipsoids (hence the name elliptical)
- Affine transformations: $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ has a representation of the form $\boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$ for $\mathbf{A}\mathbf{A}^\top = \Sigma$ and $\mathbf{Y} \sim S_k(\psi)$. So for $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{B} \in \mathbb{R}^{m\times d}$, one has $\mathbf{b} + \mathbf{B}\mathbf{X} = \mathbf{b} + \mathbf{B}(\boldsymbol{\mu} + \mathbf{A}\mathbf{Y}) \sim E_m\left(\mathbf{b} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\mathbf{A}\mathbf{A}^\top\mathbf{B}^\top, \psi\right) = E_m\left(\mathbf{b} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\Sigma\mathbf{B}^\top, \psi\right)$

  In particular, $\mathbf{a}^\top\mathbf{X} \sim E_1\left(\mathbf{a}^\top\boldsymbol{\mu}, \mathbf{a}^\top\Sigma\mathbf{a}, \psi\right)$

  By taking $\mathbf{a} = \mathbf{e}_j$, then all marginal distributions are of the same type.
- Marginals: If $\mathbf{X} = (\mathbf{Y}_1, \mathbf{Y}_2) \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$, then $\mathbf{Y}_1 \sim E_k\left(\boldsymbol{\mu}_1, \Sigma_{11}, \psi\right)$ and $\mathbf{Y}_2 \sim E_{d-k}\left(\boldsymbol{\mu}_2, \Sigma_{22}, \psi\right)$
- Conditional distributions of elliptical distributions are elliptical.
- Quadratic forms: $(\mathbf{X}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu}) = R^2$
  - If $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$, then $R^2 \sim \chi^2_d$
  - If $\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$, then $R^2/d \sim F(d, \nu)$
- Convolutions: If $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ and $\mathbf{Y} \sim E_d(\tilde{\boldsymbol{\mu}}, c\Sigma, \tilde{\psi})$ are ind, then $a\mathbf{X} + b\mathbf{Y}$ is elliptical.

Proposition (Subadditivity of VaR in elliptical models)

Let $L_i = \mathbf{v}_i^\top\mathbf{X}, \mathbf{v}_i \in \mathbb{R}^d, i = 1, \ldots, n$, where $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$.

Then $\mathrm{VaR}_\alpha\left(\sum_{i=1}^n L_i\right) \leq \sum_{i=1}^n \mathrm{VaR}_\alpha(L_i)$ for all $\alpha \in (\frac{1}{2}, 1)$. In particular, $\mathrm{VaR}_\alpha\left(\sum_{i=1}^d X_i\right) \leq \sum_{i=1}^d \mathrm{VaR}_\alpha(X_i)$ for all $\alpha \in (\frac{1}{2}, 1)$

Proof: Consider a RV of the form $L = \mathbf{v}^\top\mathbf{X} \overset{(d)}{=} \mathbf{v}^\top\boldsymbol{\mu} + \mathbf{v}^\top\mathbf{A}\mathbf{Y}$, where $\mathbf{A}\mathbf{A}^\top = \Sigma$ and $\mathbf{Y} \in \mathcal{S}_k(\psi)$.

Since $\mathbf{v}^\top\mathbf{A}\mathbf{Y} \overset{(d)}{=} \left\|\mathbf{A}^\top\mathbf{v}\right\|Y_1$, one has $L \overset{(d)}{=} \mathbf{v}^\top\boldsymbol{\mu} + \left\|\mathbf{A}^\top\mathbf{v}\right\|Y_1$, and therefore, $\mathrm{VaR}_\alpha(L) = \mathbf{v}^\top\boldsymbol{\mu} + \left\|\mathbf{A}^\top\mathbf{v}\right\|\mathrm{VaR}_\alpha(Y_1)$.
Since Y_1 is symmetric, one has $\mathrm{VaR}_\alpha(Y_1) \geq 0$ for $\alpha \in (1/2, 1)$, and hence,

$$\mathrm{VaR}_\alpha\left(\sum_{i=1}^n L_i\right) = \sum_{i=1}^n \mathbf{v}_i^\top\boldsymbol{\mu} + \left\|\sum_{i=1}^n \mathbf{A}_i^\top\mathbf{v}\right\|\mathrm{VaR}_\alpha(Y_1)$$
$$\leq \sum_{i=1}^n \mathbf{v}_i^\top\boldsymbol{\mu} + \left\|\mathbf{A}_i^\top\mathbf{v}\right\|\mathrm{VaR}_\alpha(Y_1) = \sum_{i=1}^n \mathrm{VaR}_\alpha(L_i)$$

6.4 Dimension Reduction Techniques

Factor models

Idea: Explain the variability of a d-dimensional vector X of risk factor changes with of a few underlying factors.
Def: $\mathbf{X}$ follows a p-factor model if $\mathbf{X} = \mathbf{a} + \mathbf{B}\mathbf{F} + \varepsilon$, where

(1) $\mathbf{a} \in \mathbb{R}^d$ and $\mathbf{B} \in \mathbb{R}^{d\times p}$ is a matrix of factor loadings

(2) $\mathbf{F} = \left(F_1, \ldots, F_p\right)$ is a random vec of underlying factors with $p < d$ and $\boldsymbol{\Theta} := \mathrm{Cov}(\mathbf{F})$ (systematic risk)

(3) $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_d)$ is the random vec of idiosyncratic error terms with $\mathbb{E}[\varepsilon] = \mathbf{0}$, $\boldsymbol{\Upsilon} := \mathrm{Cov}(\varepsilon)$ is diagonal and $\mathrm{Cov}(\mathbf{F}, \varepsilon) = \mathbf{0}$ (idiosyncratic risk)

- Goal: Identify or estimate $(\mathbf{a}, \mathbf{B})$ or $\mathbf{F} = \left(F_1, \ldots, F_p\right)$.
- Factor models imply that $\Sigma := \mathrm{Cov}(\mathbf{X}) = \mathbf{B}\boldsymbol{\Theta}\mathbf{B}^\top + \boldsymbol{\Upsilon}$
- For $\mathbf{B}^* = \mathbf{B}\boldsymbol{\Theta}^{1/2}$ and $\mathbf{F}^* = \boldsymbol{\Theta}^{-1/2}(\mathbf{F} - \mathbb{E}[\mathbf{F}])$, one has $\mathbf{X} = \boldsymbol{\mu} + \mathbf{B}^*\mathbf{F}^* + \varepsilon$, where $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$
- One has $\Sigma = \mathbf{B}^*(\mathbf{B}^*)^\top + \boldsymbol{\Upsilon}$, and conversely, if $\mathrm{Cov}(\mathbf{X}) = \mathbf{B}^*(\mathbf{B}^*)^\top + \boldsymbol{\Upsilon}$ for some $\mathbf{B}^* \in \mathbb{R}^{d\times p}$ with $\mathrm{rank}(\mathbf{B}^*) = p < d$ and diagonal matrix $\boldsymbol{\Upsilon}$, then $\mathbf{X}$ has a factor-model representation for a p-dimensional $\mathbf{F}$ and a d-dimenstional ε

Statistical Estimation Strategies

Consider $\mathbf{X}_t = \mathbf{a} + \mathbf{B}\mathbf{F}_t + \varepsilon_t$, $t = 1, \ldots, n$.
Three types of factor models are commonly used:

(1) Macroeconomic factor models: It is assumed that $\mathbf{F}_t, t = 1, \ldots, n$, are observable. Estimation of $\mathbf{a}$ and $\mathbf{B}$ is accomplished by time series regression

(2) Fundamental factor models: It is assumed that $\mathbf{a}$ and $\mathbf{B}$ are known but the factors $\mathbf{F}_t$ are unobserved and have to be estimated from $\mathbf{X}_t, t = 1, \ldots, n$, using cross-sectional regression at each t.

(3) Statistical factor models: It is assumed that neither $(\mathbf{a}, \mathbf{B})$ nor the factors $\mathbf{F}_t$ are observed (both have to be estimated from $\mathbf{X}_t, t = 1, \ldots, n$). The factors can be found with principal component analysis.

Estimating Macroeconomic Factor Models

Univariate Regression

- Consider the (univariate) time series regression model

  $$\mathbf{X}_{t,j} = \mathbf{a}_j + \mathbf{b}_j^\top\mathbf{F}_t + \varepsilon_{t,j}, \quad t = 1, \ldots, n$$
- The components $\mathbf{F}_{t,1}, \ldots, \mathbf{F}_{t,p}$ are assumed to be observable changes in macroeconomic factors, such as index returns, interest rates, inflation, GDP growth, unemployment rate, ...
- To justify the use of ordinary least-squares (OLS) to derive statistical properties of the method it is usually assumed that, conditional on the factors, the errors $\varepsilon_{1,j}, \ldots, \varepsilon_{n,j}$ form a white noise process (i.e. are identically distributed and serially uncorrelated)
- $\hat{\mathbf{a}}_j$ estimates $\mathbf{a}_j$, $\hat{\mathbf{b}}_j$ estimates the j-th row of $\mathbf{B}$
- Models can also be estimated simultaneously using multivariate regression.

Estimating fundamental factor models

- Consider the cross-sectional regression model $\mathbf{X}_t = \mathbf{B}\mathbf{F}_t + \varepsilon_t$ ($\mathbf{B}$ is assumed to be known; $\mathbf{F}_t$ to be estimated; $\mathrm{Cov}(\varepsilon) = \boldsymbol{\Upsilon}$)
  Note that $\mathbf{a}$ can be absorbed into $\mathbf{F}_t$.
  To obtain precision in estimating $\mathbf{F}_t$, one needs $p \ll d$
- E.g. it is assumed that stock returns of companies in the same country/industry are affected by a common factor
- First estimate $\mathbf{F}_t$ via OLS by $\hat{\mathbf{F}}_t^{\mathrm{OLS}} = \left(\mathbf{B}^\top\mathbf{B}\right)^{-1}\mathbf{B}^\top\mathbf{X}_t$. This is the best linear unbiased estimator if $\mathrm{Cov}(\varepsilon_t) = \sigma^2\mathbf{I}_d$ for some $\sigma > 0$

- However, if $\mathrm{Cov}(\varepsilon_t) = \Upsilon$ for a general $d \times d$ diagonal matrix Υ, it is possible to obtain linear unbiased estimates with smaller squared errors via **generalized least squares (GLS)**
- To do that, estimate Υ by $\hat{\Upsilon}$ via the diagonal of the sample covariance matrix of the residuals $\hat{\varepsilon}_t = \mathbf{X}_t - \mathbf{B}\hat{\mathbf{F}}_t^{\mathrm{OLS}}$, $t = 1, \ldots, n$
- Then estimate $\mathbf{F}_t$ by $\hat{\mathbf{F}}_t = \left(\mathbf{B}^\top \hat{\Upsilon}^{-1} \mathbf{B}\right)^{-1} \mathbf{B}^\top \hat{\Upsilon}^{-1} \mathbf{X}_t$

Estimating statistical factor models with principal component analysis (PCA)

- **Goal**: Reduce the dimensionality of highly correlated data by finding a small number of uncorrelated linear combinations which account for most of the variance in the data; this can be used for finding factors
- **Key**: Every symmetric matrix $\mathbf{M}$ admits a spectral decomposition $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where (1) $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ is the diagonal matrix of eigenvalues of $\mathbf{M}$, which w.l.o.g. are ordered so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$, and (2) $\mathbf{U}$ is an orthogonal matrix whose columns are eigenvectors of $\mathbf{M}$ of length 1.
- Let $\Sigma = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ with $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0$ and $\mathbf{Y} = \mathbf{U}^\top(\mathbf{X} - \boldsymbol{\mu})$ the **principle component transform** $Y_j = \mathbf{u}_j^\top(\mathbf{X} - \boldsymbol{\mu})$ is the j-th principle component of $\mathbf{X}$ (where $\mathbf{u}_j$ is the j-th column of $\mathbf{U}$)
- One has $\mathbb{E}\mathbf{Y} = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{Y}) = \mathbb{E}\left[\mathbf{Y}\mathbf{Y}^\top\right] = \mathbf{U}^\top \Sigma \mathbf{U} = \mathbf{\Lambda}$, so the principal components are uncorrelated and $\mathrm{Var}\left(Y_j\right) = \lambda_j$.

  The principal components are thus ordered by decreasing variance.
- One can show (1) The first principal component is that normalized linear combination of $\mathbf{X}$ which has maximal variance among all such combinations, i.e. $\mathrm{Var}\left(\mathbf{u}_1^\top \mathbf{X}\right) = \max\left\{\mathrm{Var}\left(\mathbf{u}^\top \mathbf{X}\right) : \mathbf{u}^\top \mathbf{u} = 1\right\}$;
  (2) For $j = 2, \ldots, d$, the j-th principal component is that normalized linear combination of $\mathbf{X}$ which has maximal variance among all such linear combinations which are orthogonal to (and hence uncorrelated with) the first $j-1$ principal components.
- $\sum_{j=1}^d \mathrm{Var}\left(Y_j\right) = \sum_{j=1}^d \lambda_j = \sum_{j=1}^d \mathrm{Var}\left(X_j\right)$. So $\sum_{j=1}^k \lambda_j / \sum_{j=1}^d \lambda_j$ can be interpreted as the fraction of total variance explained by the first k principal components

Principal components as factors

- Inverting the principal component transform $\mathbf{Y} = \mathbf{U}^\top(\mathbf{X} - \boldsymbol{\mu})$, one obtains

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{U}\mathbf{Y} = \boldsymbol{\mu} + \mathbf{U}'\mathbf{Y}' + \mathbf{U}''\mathbf{Y}'' =: \boldsymbol{\mu} + \mathbf{U}'\mathbf{Y}' + \varepsilon$$

  where $\mathbf{Y}' \in \mathbb{R}^k$ contains the first k principal components. This is reminiscent of the basic factor model.
- Although $\varepsilon_1, \ldots, \varepsilon_d$ will tend to have small variances, the assumptions of the factor model are generally violated (since they need not have a diagonal covariance matrix and need not be uncorrelated with $\mathbf{Y}'$). Nevertheless, principal components are often interpreted as factors.
- The same can be applied to the sample covariance matrix to obtain the **sample principal components**.

7 Copulas and Dependence

7.1 Copulas

Advantages

- Most natural in a static distributional context (no time dependence; apply, e.g. to residuals of an ARMA-GARCH model)
- Copulas allow one to understand and study dependence independently of the margins (first part of Sklar's theorem)
- Copulas allow for a bottom-up approach to multivariate model building by combining marginal distributions with a given dependence structure (second part of Sklar's theorem)

Definition (Copula)

A **copula** C is a multivariate cdf with $\mathrm{Unif}(0,1)$ margins.

Characterization

A mapping $C : [0,1]^d \to [0,1]$ is a copula **iff**

(1) C is **grounded**, that is, $C(u_1, \ldots, u_d) = 0$ if $u_j = 0$ for at least one $j \in \{1, \ldots, d\}$

(2) C has **standard uniform** one-dim marginals, that is, $C\left(1, \ldots, 1, u_j, 1, \ldots, 1\right) = u_j$ for all $u_j \in [0,1]$ and $j \in \{1, \ldots, d\}$

(3) C is d-**monotone**, that is, for all $a, b \in [0,1]^d$ such that $a \leq b$, $\Delta_{(a,b]} C = \sum_{i \in \{0,1\}^d} (-1)^{\sum_{j=1}^d i_j} C\left(a_1^{i_1} b_1^{1-i_1}, \ldots, a_d^{i_d} b_d^{1-i_d}\right) \geq 0$

Proof: It is clear that a copula satisfies (1)-(3).

On the other hand, if a function $C : [0,1]^d \to [0,1]$ satisfies (1)-(3), then for $a, b \in [0,1]^d$, $|C(b) - C(a)|$

$$\leq \sum_{j=1}^d \left|C\left(b_1, \ldots, b_j, a_{j+1}, \ldots, a_d\right) - C\left(b_1, \ldots, b_{j-1}, a_j, \ldots, a_d\right)\right|$$

$$\leq \sum_{j=1}^d \left|C\left(1, \ldots, 1, b_j, 1, \ldots, 1\right) - C\left(1, \ldots, 1, a_j, 1, \ldots, 1\right)\right|$$

$$= \sum_{j=1}^d \left|b_j - a_j\right|$$

First $\leq$ by Δ-inequality; Second $\leq$ by d-monotonicity; Third $=$ by uniform marginals.

In particular, C is continuous, and hence, fulfills all the properties of a d-dimensional cdf with uniform marginals

Lemma (Quantile transformation)

Let $q : (0,1) \to \mathbb{R}$ be a quantile func of a RV X and $U \sim \mathrm{Unif}(0,1)$
(that is, $\mathbb{P}[X < q(u)] \leq u \leq \mathbb{P}[X \leq q(u)]$ for all $u \in (0,1)$).
Then $q(U)$ has the same dist as X.

Proof: $q(u) \leq x$ implies $u \leq F_X(x)$ and $u < F_X(x)$ implies $q(u) \leq x$. It follows that $\mathbb{P}[q(U) \leq x] = \mathbb{P}[U \leq F_X(x)] = F_X(x)$.

Lemma (Probability transformation)

Let X be a RV with continuous cdf F_X. Then $F_X(X) \sim \mathrm{Unif}(0,1)$.

Proof: Let q be a quantile function of X. Since F_X is continuous, one has

$$\mathbb{P}[F_X(X) \leq u] = \mathbb{P}[X \leq q(u)] = F_X(q(u)) = u$$

for all $u \in (0,1)$.

Note that if F_X **not continuous**, its image is contained in $[0,1] \backslash I$ for a non-empty interval I. So $F_X(X)$ cannot be $\mathrm{Unif}(0,1)$.

Sklar's Thm

(1) For any d-dimensional cdf F with marginals $F_1, \ldots, F_d$, there exists a copula C such that $F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)), x \in \mathbb{R}^d$ (*).
If $F_1, \ldots, F_d$ are continuous, then C is **unique** and given by $C(u_1, \ldots, u_d) = F(q_1(u_1), \ldots, q_d(u_d)), u \in (0,1)^d$ where $q_1, \ldots, q_d$ are (arbitrary) quantile functions of $F_1, \ldots F_d$.

(2) Conversely, given a d-dimensional copula C and one-dimensional cdf's $F_1, \ldots, F_d$, (*) defines a d-dim cdf with one-dim marginals $F_1, \ldots, F_d$.

Proof of Sklar's Thm

(1) For simplicity, we assume for the proof of this direction that $F_1, \ldots, F_d$ are continuous.

Let $\mathbf{X} \sim F$ and set $U_j = F_j\left(X_j\right) \sim \mathrm{Unif}(0,1), j = 1, \ldots, d$. So the cdf C of U is a copula.
Moreover, let $q_1, \ldots, q_d$ be quantile functions of $X_1, \ldots, X_d$.
Then $X_j \overset{\text{a.s.}}{=} q_j\left(F_j\left(X_j\right)\right) = q_j\left(U_j\right)$, therefore,

$$\begin{aligned} F(\mathbf{x}) &= \mathbb{P}\left[X_j \leq x_j, j = 1, \ldots, d\right] \\ &= \mathbb{P}\left[q_j\left(U_j\right) \leq x_j, j = 1, \ldots, d\right] \\ &= \mathbb{P}\left[U_j \leq F_j\left(x_j\right), j = 1, \ldots, d\right] \\ &= C(F_1(x_1), \ldots, F_d(x_d)), \quad \mathbf{x} \in \mathbb{R}^d \end{aligned}$$

Hence, (*) holds.
In addition, since F_j is continuous, one has $F_j\left(q_j\left(u_j\right)\right) = u_j$ for all $u_j \in (0,1)$. So

$$\begin{aligned} C(u_1, \ldots, u_d) &= C(F_1(q_1(u_1)), \ldots, F_d(q_d(u_d))) \\ &= F(q_1(u_1), \ldots, q_d(u_d)), \mathbf{u} \in (0,1)^d \end{aligned}$$

(2) Let $U \sim C$, and let $q_1, \ldots, q_d$ be quantile functions of $F_1, \ldots, F_d$.
Define $\mathbf{X} = (q_1(U_1), \ldots, q_d(U_d))$. Then

$$\begin{aligned} \mathbb{P}[\mathbf{X} \leq \mathbf{x}] &= \mathbb{P}\left[q_j\left(U_j\right) \leq x_j, j = 1, \ldots, d\right] \\ &= \mathbb{P}\left[U_j \leq F_j\left(x_j\right), j = 1, \ldots, d\right] \\ &= C(F_1(x_1), \ldots, F_d(x_d)), \quad \mathbf{x} \in \mathbb{R}^d \end{aligned}$$

So $F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$ is the cdf of X, and the marginals of F are $F_1, \ldots, F_d$

Example: Bivariate Bernoulli distribution

Let (X_1, X_2) follow a bivariate Bernoulli distribution with $\mathbb{P}[X_1 = k, X_2 = l] = 1/4, k, l \in \{0,1\}$. Then $\mathbb{P}\left[X_j = k\right] = 1/2, k = 0, 1$, and $\mathrm{Im}\left(F_j\right) = \{0, 1/2, 1\}, j = 1, 2$.
Any copula with $C(1/2, 1/2) = 1/4$ satisfies $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$, $(x_1, x_2) \in \mathbb{R}^2$;
e.g. the **independence copula** $C(u_1, u_2) = u_1 u_2$ or the **diagonal copula** $\min\left\{u_1, u_2, \left(u_1^2 + u_2^2\right)/2\right\}$

- $\mathbf{X}$ (or F) with margins $F_1, \ldots, F_d$ has copula C if $F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d))$
- A **copula model** for $\mathbf{X}$ means $F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d))$ for some (parametric) copula C and (parametric) marginals $F_1, \ldots, F_d$.

Corollary

Let $\mathbf{X}$ be a random vec such that all X_j have continuous cdf $F_j, j = 1, \ldots, d$. Then $\mathbf{X}$ has copula $C \Leftrightarrow (F_1(X_1), \ldots, F_d(X_d))$ has cdf C

Thm (Invariance principle)

Let $\mathbf{X}$ be a random vec with continuous margins $F_1, \ldots, F_d$ and copula C. If $T_j : \mathrm{Im}\left(X_j\right) \to \mathbb{R}$, $j = 1, \ldots, d$, are strictly increasing, then $(T_1(X_1), \ldots, T_d(X_d))$ also has copula C.

Proof: Since T_j is strictly increasing, $T_j\left(X_j\right)$ is continuously distributed. So for $x \in \mathrm{Im}\left(T_j\right)$, $F_{T_j(X_j)}(x) = \mathbb{P}\left[T_j\left(X_j\right) \leq x\right] = \mathbb{P}\left[X_j \leq T_j^{-1}(x)\right] = F_j\left(T_j^{-1}(x)\right)$, where T_j^{-1} is the generalized inverse.

Hence, $\mathbb{P}\left[F_{T_j(X_j)}\left(T_j\left(X_j\right)\right) \leq u_j \text{ for all } j\right] = \mathbb{P}\left[F_j\left(T_j^{-1}\left(T_j\left(X_j\right)\right)\right) \leq u_j \text{ for all } j\right] = \mathbb{P}\left[F_j\left(X_j\right) \leq u_j \text{ for all } j\right] = C(\mathbf{u})$

Interpretation of Sklar's Thm and the Invariance Principle

- Part 1 of Sklar's theorem allows one to **decompose any cdf 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** $U = (F_1(X_1), \ldots, F_d(X_d))$ instead of $X = (X_1, \ldots, X_d)$ (they both have the same copula!). This is interesting for statistical applications, e.g. parameter estimation or goodness-of-fit tests

- Part 2 allows one to construct flexible multivariate distributions for particular applications

Fréchet–Hoeffding bounds

(1) For any d-dimensional copula C,

$$W(u) \le C(u) \le M(u), u \in [0,1]^d$$

where $W(u) = \left(\sum_{j=1}^d u_j - d + 1\right)^+$ and $M(u) = \min_{1\le j\le d} u_j$

(2) W is a copula **iif** $d = 2$

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

Proof: $C(u) \le C\left(1,\dots,1,u_j,1,\dots,1\right) = u_j$ for all j. So $C(u) \le M(u), u \in [0,1]^d$.
If $U \sim \mathrm{Unif}(0,1)$, then $\mathbb{P}[U \le u_1,\dots,U \le u_d] = \min_j u_j$. So $(U,\dots,U)$ has copula M.
$1 - C(u) = C(1,\dots,1) - C(u) \le \sum_{j=1}^d \left(1-u_j\right) = d - \sum_{j=1}^d u_j$. So $C(u) \ge W(u)$.
For $d = 2, (U, 1-U)$ has cdf W.
For $d \ge 3, W$ violates d-monotonicity. So it cannot be a copula.

- The Fréchet-Hoeffding bounds correspond to perfect dependence (**negative for W; positive for M**)
- The Fréchet-Hoeffding bounds lead to bounds for any cdf $F : \mathbb{R}^d \to [0,1]$:

$$\left(\sum_{j=1}^d F_j\left(x_j\right) - d + 1\right)^+ \le F(\mathbf{x}) = C\left(F_1\left(x_1\right),\dots,F_d\left(x_d\right)\right) \le \min_{1\le j\le d} F_j\left(x_j\right)$$

- The Fréchet-Hoeffding bound M is the comonotonicity copula. It is the cdf of $(U,\dots,U)$. If the copula of $(X_1,\dots,X_d)$ is M, then $(X_1,\dots,X_d) \overset{(d)}{=} (q_1(U),\dots,q_d(U))$ for $U \sim \mathrm{Unif}(0,1)$ and quantile functions q_j of F_j. We say $X_1,\dots,X_d$ are comonotonic or perfectly positively dependent.
- For $d = 2$, the Fréchet-Hoeffding bound W is the counter-monotonicity copula. It is the cdf of $(U, 1-U)$. If the copula of (X_1, X_2) is W, then $(X_1, X_2) \overset{(d)}{=} (q_1(U), q_2(1-U))$ for $U \sim \mathrm{Unif}(0,1)$ and quantile functions q_j of F_j. We say X_1 and X_2 are counter-monotonic or perfectly neg dep.

cdf's with densities have copulas with densities

- Let F be a d-dimensional cdf with density f and copula C.
- Then $F(x_1,\dots,x_d) = C\left(F_1(x_1),\dots,F_d(x_d)\right)$ and so $C(u_1,\dots,u_d) = F\left(q_1(u_1),\dots,q_d(u_d)\right)$, where q_j are quantile functions of F_j
- $$q_j'\left(u_j\right) = \frac{1}{F_j'\left(q_j\left(u_j\right)\right)} = \frac{1}{f_j\left(q_j\left(u_j\right)\right)}$$
  for almost all $u_j \in [0,1]$.
- Therefore, C has a density c, given by

$$\begin{aligned} c(\mathbf{u}) &= \frac{\partial}{\partial u_1}\cdots\frac{\partial}{\partial u_d} C(u_1,\dots,u_d) \\ &= \frac{\partial}{\partial x_1}\cdots\frac{\partial}{\partial x_d} F\left(q_1(u_1),\dots,q_d(u_d)\right)\prod_{j=1}^d q_j'\left(u_j\right) \\ &= \frac{f\left(q_1(u_1),\dots,q_d(u_d)\right)}{\prod_{j=1}^d f_j\left(q_j\left(u_j\right)\right)} \end{aligned}$$

where f_j is the density of F_j.

Examples of Copulas

(i) Fundamental Copulas

- $\Pi(\mathbf{u}) = \prod_{j=1}^d u_j$ is the independence copula since $F(\mathbf{x}) = \Pi\left(F_1(x_1),\dots,F_d(x_d)\right)$ implies $F(\mathbf{x}) = \prod_{j=1}^d F_j\left(x_j\right)$. So $X_1,\dots,X_d$ are ind if copula is Π.
- The Fréchet–Hoeffding bound M is the comonotonicity copula.
- For $d = 2$, the Fréchet-Hoeffding bound W is the counter-monotonicity copula.

(ii) Implicit Copulas

Elliptical copulas are implicit copulas arising from elliptical distributions via Sklar's theorem. The two most prominent parametric families in this class are Gauss copulas and t-copulas.

Gauss copulas

- Let $\mathbf{X} \sim \mathcal{N}_d(\mathbf{0},\mathbf{P})$, where $\mathbf{P}$ is a $d \times d$ correlation matrix.
  The corresponding Gauss copula is $C_{\mathbf{P}}^{Ga}(\mathbf{u}) = \Phi_{\mathbf{P}}\left(\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_d)\right)$, where Φ is the cdf of $\mathcal{N}(0,1)$ and $\Phi_{\mathbf{P}}$ the cdf of $\mathcal{N}_d(0,\mathbf{P})$
- If $\mathbf{P} = \mathbf{I}_d$, then $C = \Pi$ (independence copula)
- If $\mathbf{P} = \mathbf{1}\mathbf{1}^\top$, then $C = M$ (comonotonicity copula)
- If $d = 2$ and $\rho = \mathbf{P}_{12} = -1$, then $C = W$ (counter-monotonicity copula)
- The density $c_{\mathbf{P}}^{Ga}$ of $C_{\mathbf{P}}^{Ga}$ is given by

$$\begin{aligned} c_{\mathbf{P}}^{Ga}(\mathbf{u}) &= \frac{\varphi_{\mathbf{P}}\left(\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_d)\right)}{\prod_{j=1}^d \varphi\left(\Phi^{-1}\left(u_j\right)\right)} \\ &= \frac{1}{\sqrt{\det \mathbf{P}}}\exp\left(-\frac{1}{2}\mathbf{x}^\top\left(\mathbf{P}^{-1}-\mathbf{I}_d\right)\mathbf{x}\right) \end{aligned}$$

where $\mathbf{x} = \left(\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_d)\right)$

t-copulas

- Let $\mathbf{X} \sim t_d(\nu,\mathbf{0},\mathbf{P})$ for a parameter $\nu > 0$ and correlation matrix $\mathbf{P}$. The corresponding t-copula is $C_{\nu,\mathbf{P}}^t(\mathbf{u}) = t_{\nu,\mathbf{P}}\left(t_\nu^{-1}(u_1),\dots,t_\nu^{-1}(u_d)\right)$, where $t_{\nu,\mathbf{P}}$ is the cdf of $t_d(\nu,\mathbf{0},\mathbf{P})$ and t_ν the cdf of the univariate t_ν-distribution.
- If $\mathbf{P} = \mathbf{1}\mathbf{1}^\top$, then $C = M$ (comonotonicity copula)
- If $d = 2$ and $\mathbf{P}_{12} = -1$, then $C = W$ (counter-monotonicity copula)
- However, if $\mathbf{P} = \mathbf{I}_d$, it does not follow that $C = \Pi$ (unless $\nu = \infty$, in which case $C_{\nu,\mathbf{P}}^t = C_{\mathbf{P}}^{Ga}$)
- The density $c_{\mu,P}^t$ of $C_{\nu,P}^t$ is given by

$$c_{\nu,P}^t(\mathbf{u}) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\det \mathbf{P}}}\left(\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right)^d \frac{\left(1+\mathbf{x}^\top\mathbf{P}^{-1}\mathbf{x}/\nu\right)^{-(\nu+d)/2}}{\prod_{j=1}^d\left(1+x_j^2/\nu\right)^{-(\nu+1)/2}}$$

for $\mathbf{x} = \left(t_\nu^{-1}(u_1),\dots,t_\nu^{-1}(u_d)\right)$

Advantages and drawbacks of elliptical copulas

Advantages: Flexible class for modeling dependencies; Densities available; Sampling (typically) simple.
Drawbacks: Typically, C is not explicit; Radially symmetric (so the same lower/upper tail behavior)

(iii) Explicit Copulas

- Archimedean copulas are copulas of the form $C(\mathbf{u}) = \psi\left(\psi^{-1}(u_1)+\cdots+\psi^{-1}(u_d)\right), \mathbf{u} \in [0,1]^d$, for a generator $\psi : [0,\infty) \to [0,1]$ satisfying (1) $\psi(0) = 1$, (2) $\lim_{x\to\infty}\psi(x) = 0$, (3) ψ is continuous, non-increasing and strictly decreasing on $[0, \inf\{x : \psi(x) = 0\}]$
- We denote the set of all generators by Ψ
- If $\psi(x) > 0$ for all $x \in [0,\infty)$, we call ψ strict
- We set $\psi^{-1}(0) = \inf\{x : \psi(x) = 0\}$

Clayton copula: $\psi(x) = (1+x)^{-1/\theta}$ for a parameter $\theta \in (0,\infty)$, $C_\theta^C(\mathbf{u}) = \left(u_1^{-\theta}+\cdots+u_d^{-\theta}-d+1\right)^{-1/\theta}$.
For $\theta \downarrow 0, C \to \Pi$; For $\theta \uparrow \infty, C \to M$.
Gumbel copula: $\psi(x) = \exp\left(-x^{1/\theta}\right)$ for a parameter $\theta \in [1,\infty)$, $C_\theta^{Gu}(\mathbf{u}) = \exp\left(-\left((-\log u_1)^\theta+\cdots+(-\log u_d)^\theta\right)^{1/\theta}\right)$.
For $\theta = 1, C = \Pi$; For $\theta \to \infty, C \to M$.

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 (exchangeability)
- Often used only with a small number of parameters (some extensions available)

Meta distributions

- Fréchet class: Class of all cdf's F with given marginals $F_1,\dots,F_d$.
- Meta-C models: All cdf's F with the same given copula C.
- Example: A meta-Gauss model consists of cdf's F with a given Gauss copula C_P^{Ga} and some marginals $F_1,\dots,F_d$

Sampling implicit copulas

(1) Sample $\mathbf{X} \sim F$, where F is a cdf with continuous margins $F_1,\dots,F_d$

(2) Return $\mathbf{U} = \left(F_1(X_1),\dots,F_d(X_d)\right)$ (probability transformation)

Examples

- Sampling Gauss copulas C_P^{Ga}: (1) Sample $\mathbf{X} \sim \mathcal{N}_d(0,\mathbf{P})$, $\mathbf{X} \overset{(d)}{=} \mathbf{A}\mathbf{Z}$ for $\mathbf{A}\mathbf{A}^\top = \mathbf{P}, \mathbf{Z} \sim \mathcal{N}_d(0,\mathbf{I}_d)$; (2) Return $\mathbf{U} = \left(\Phi(X_1),\dots,\Phi(X_d)\right)$
- Sampling t-copulas $C_{\nu,P}^t$: (1) Sample $\mathbf{X} \sim t_d(\nu,\mathbf{0},\mathbf{P})$, $\mathbf{X} \overset{(d)}{=} \sqrt{W}\mathbf{A}\mathbf{Z}$ for $W = 1/G, G \sim \mathrm{Gamma}(\nu/2,\nu/2)$; (2) Return $\mathbf{U} = \left(t_\nu(X_1),\dots,t_\nu(X_d)\right)$

Sampling meta distributions

(1) Sample $\mathbf{U} \sim C$

(2) Return $\mathbf{X} = \left(q_1(U_1),\dots,q_d(U_d)\right)$, where q_j are quantile funcs of F_j (quantile transformation)

Survival copulas

- If $\mathbf{U} \sim C$, then $1 - \mathbf{U} \sim \hat{C}$. $\hat{C}$ is the survival copula of C.
- $\hat{C}$ can be expressed as

$$\hat{C}(\mathbf{u}) = \sum_{J\subseteq\{1,\dots,d\}} (-1)^{|J|} C\left((1-u_1)^{I_J(1)},\dots,(1-u_d)^{I_J(d)}\right)$$

in terms of its corresponding copula (essentially an application of the Poincaré-Sylvester sieve formula).

For $d=2$,

$$\begin{aligned}&\hat{C}(u_1,u_2)\\=&1-(1-u_1)-(1-u_2)+C(1-u_1,1-u_2)\\=&-1+u_1+u_2+C(1-u_1,1-u_2)\end{aligned}$$

- If C has a density c, density of $\hat{C}$ is $\hat{c}(u)=c(1-u)$
- If $\hat{C}=C, C$ is called radially symmetric.
  Note: Π, M and W are all radially symmetric.
- If X_j is symmetrically distributed around $a_j, j\in\{1,\dots,d\}$, then $\mathbf{X}$ is radially symmetric around a if and only if $C=\hat{C}$
- Sklar's thm can also be formulated for survival funcs. In this case, $\bar{F}(\mathbf{x})=\hat{C}\left(\bar{F}_1(x_1),\dots,\bar{F}_d(x_d)\right)$, where $\bar{F}(\mathbf{x})=\mathbb{P}[\mathbf{X}>\mathbf{x}]$ and $\bar{F}_j(x)=\mathbb{P}\left[X_j>x\right]$.
  Survival copulas combine marginal survival functions with joint survival functions.
  Note that $\hat{C}$ is a cdf, but $\bar{F}$ and $\bar{F}_1,\dots,\bar{F}_d$ are not!

Copula densities

- By Sklar's theorem, $F(\mathbf{x})=C(F_1(x_1),\dots,F_d(x_d))$. So if all F_j have densities f_j and C has a density c, then F has a density of the form

$$\begin{aligned}f(\mathbf{x})&=\frac{\partial}{\partial u_1}\cdots\frac{\partial}{\partial u_d}C(F_1(x_1),\dots,F_d(x_d))\prod_{j=1}^d F_j'(x_j)\\&=c(F_1(x_1),\dots,F_d(x_d))\prod_{j=1}^d f_j(x_j)\end{aligned}$$

This gives the following formula for c :

$$c(\mathbf{u})=\frac{f(q_1(u_1),\dots,q_d(u_d))}{f_1(q_1(u_1))\cdots f_d(q_d(u_d))}.$$

- Moreover, it follows that the log-density splits into $\log f(\mathbf{x})=\log c(F_1(x_1),\dots,F_d(x_d))+\sum_{j=1}^d\log f_j(x_j)$, which allows for a two-stage estimation (marginal and copula parameters).

Exchangeability

- $\mathbf{X}$ is exchangeable if $(X_1,\dots,X_d)\overset{(d)}{=}\left(X_{\pi(1)},\dots,X_{\pi(d)}\right)$ for any permutation $(\pi(1),\dots,\pi(d))$ of $(1,\dots,d)$
- A copula C is exchangeable if it is the cdf of an exchangeable U with Unif(0,1) margins. This holds **iff** $C(u_1,\dots,u_d)=C\left(u_{\pi(1)},\dots,u_{\pi(d)}\right)$ for all possible permutations of $(1,\dots,d)$, that is, if C is symmetric.
- Exchangeable/symmetric copulas are useful for modeling homogeneous portfolios.
- **Examples**: Archimedean copulas; Elliptical copulas (such as Gauss/t) for equicorrelated $\mathbf{P}$, i.e. $\mathbf{P}=\rho\mathbf{1}\mathbf{1}^\top+(1-\rho)\mathbf{I}_d$ for $\rho\geq-1/(d-1)$

7.2 Dependence concepts and measures

Dependence measures are scalar measures which summarize the dependence in terms of a single number.

Examples: Linear correlation; Rank correlation (Kendall's tau, Spearman's rho); Tail dependence

Perfect Dependence

$X_1,\dots,X_d$ are comonotone if $(X_1,\dots,X_d)$ has copula M. X_1,X_2 are counter-monotone if (X_1,X_2) has copula W.

Proposition (Perfect dependence)

: Let $\mathbf{X}=(X_1,\dots,X_d)$ be a random vector, $q_1,\dots,q_d$ quantile functions of the marginals and $U\sim$ Unif(0,1). Then (1) $\mathbf{X}$ has comonotonicity copula $M\Leftrightarrow X\overset{(d)}{=}(q_1(U),\dots,q_d(U))$.
(2) $d=2$ and $\mathbf{X}$ has counter-monotonicity copula $W\Leftrightarrow\mathbf{X}\overset{(d)}{=}(q_1(U),q_2(1-U))$.

Proof:

(1) $\mathbb{P}[q_1(U)\leq x_1,\dots,q_d(U)\leq x_d]=\mathbb{P}[U\leq F_1(x_1),\dots,U\leq F_d(x_d)]=\min_{1\leq j\leq d}F_j(x_j)=M(F_1(x_1),\dots,F_d(x_d))$

So X has copula $M\Leftrightarrow X\overset{(d)}{=}(q_1(U),\dots,q_d(U))$

(2) $\mathbb{P}[q_1(U)\leq x_1,q_2(1-U)\leq x_2]=\mathbb{P}[U\leq F_1(x_1),1-U\leq F_2(x_2)]=\mathbb{P}[1-F_2(x_2)\leq U\leq F_1(x_1)]=(F_1(x_1)+F_2(x_2)-1)^+=W(F_1(x_1),F_2(x_2))$

So (X_1,X_2) has copula $W\Leftrightarrow(X_1,X_2)\overset{(d)}{=}(q_1(U),q_2(1-U))$

Proposition (Comonotone additivity)

Let $X_j\sim F_j, j=1,\dots,d$, be comonotone. Then $\mathrm{VaR}_\alpha(X_1+\cdots+X_d)=\mathrm{VaR}_\alpha(X_1)+\cdots+\mathrm{VaR}_\alpha(X_d)$ for all $\alpha\in(0,1)$, and as a consequence, $\mathrm{AVaR}_\alpha(X_1+\cdots+X_d)=\mathrm{AVaR}_\alpha(X_1)+\cdots+\mathrm{AVaR}_\alpha(X_d)$ for all $\alpha\in(0,1)$.

Proof: $\mathbf{X}$ has the same distribution as $(q_1(U),\dots,q_d(U))$, where $U\sim$ Unif(0,1) and $q_1,\dots,q_d$ are left-continuous quantile functions of $X_1,\dots,X_d$.
So $X_1+\cdots+X_d$ has the same distribution as $q_1(U)+\cdots+q_d(U)$.
It follows that $\mathrm{VaR}_\alpha(X_1+\cdots+X_d)=q_1(\alpha)+\cdots+q_d(\alpha)=\mathrm{VaR}_\alpha(X_1)+\cdots+\mathrm{VaR}_\alpha(X_d)$ and $\mathrm{AVaR}_\alpha(X_1+\cdots+X_d)=\frac{1}{1-\alpha}\int_\alpha^1\sum_{j=1}^d\mathrm{VaR}_u(X_j)du=\sum_{j=1}^d\mathrm{AVaR}_\alpha(X_j)$

Linear Correlation

Def: For two RVs X_1 and X_2 satisfying $\mathbb{E}\left[X_j^2\right]<\infty$ and $\mathrm{Var}\left(X_j\right)>0, j=1,2$, the (linear or Pearson's) correlation coefficient ρ is defined by $\rho(X_1,X_2)=\frac{\mathrm{Cov}(X_1,X_2)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}}=\frac{\mathbb{E}[(X_1-\mathbb{E}X_1)(X_2-\mathbb{E}X_2)]}{\sqrt{\mathbb{E}\left[(X_1-\mathbb{E}X_1)^2\right]\mathbb{E}\left[(X_2-\mathbb{E}X_2)^2\right]}}$.

Proposition (Hoeffding's identity)

Let $X_j\sim F_j, j=1,2$, be two RVs with joint cdf F that $\mathbb{E}\left[X_j^2\right]<\infty, j=1,2$. Then

$$\mathrm{Cov}(X_1,X_2)=\iint_{\mathbb{R}^2}[F(x_1,x_2)-F_1(x_1)F_2(x_2)]dx_1dx_2$$

.

Proof: Let $\left(X_1',X_2'\right)$ be an independent copy of (X_1,X_2). Then

$$\begin{aligned}&2\mathrm{Cov}(X_1,X_2)\\=&\mathbb{E}[(X_1-\mathbb{E}X_1)(X_2-\mathbb{E}X_2)]\\&+\mathbb{E}\left[\left(X_1'-\mathbb{E}X_1'\right)\left(X_2'-\mathbb{E}X_2'\right)\right]\\=&\mathbb{E}\left[\left((X_1-\mathbb{E}X_1)-\left(X_1'-\mathbb{E}X_1'\right)\right)\right.\\&\left.\left((X_2-\mathbb{E}X_2)-\left(X_2'-\mathbb{E}X_2'\right)\right)\right]\\=&\mathbb{E}\left[\left(X_1-X_1'\right)\left(X_2-X_2'\right)\right]\\=&\mathbb{E}\left[\int_{\mathbb{R}}\int_{\mathbb{R}}\left(1_{\{X_1'\leq x_1\}}-1_{\{X_1\leq x_1\}}\right)\right.\\&\left.\left(1_{\{X_2'\leq x_2\}}-1_{\{X_2\leq x_2\}}\right)dx_1dx_2\right]\\=&\int_{\mathbb{R}}\int_{\mathbb{R}}\mathbb{E}\left[\left(1_{\{X_1'\leq x_1\}}-1_{\{X_1\leq x_1\}}\right)\left(1_{\{X_2'\leq x_2\}}-1_{\{X_2\leq x_2\}}\right)\right]dx_1dx_2\\=&2\int_{\mathbb{R}}\int_{\mathbb{R}}[F(x_1,x_2)-F_1(x_1)F_2(x_2)]dx_1dx_2\end{aligned}$$

Properties and Drawbacks of Linear Correlation

Let X_1 and X_2 be two random variables such that $\mathbb{E}\left[X_j^2\right]<\infty$ and $\mathrm{Var}\left(X_j\right)>0, j=1,2$.

- Note that ρ depends on the marginal distributions! In particular, second moments have to exist which is not the case, e.g. for $X_1,X_2\sim F(x)=\left(1-x^{-2}\right)^+$
- $|\rho|\leq1$. Furthermore, $|\rho|=1$ if and only if there exist constants $a\in\mathbb{R}\backslash\{0\}, b\in\mathbb{R}$ such that $X_2=aX_1+b$ a.s. This does not cover other strong functional dependence such as e.g. $X_2=X_1^2$
- If X_1 and X_2 are independent, then $\rho=0$. However, the converse is not true in general!
- ρ is invariant under strictly increasing linear transformations but not invariant under strictly increasing transformations in general! For instance, if $(X_1,X_2)\sim\mathcal{N}_2(0,\mathbf{P})$ for a 2×2-correlation matrix $\mathbf{P}$ with $P_{12}=\rho$, then $\rho(X_1,X_2)=\rho$ but $\rho(F_1(X_1),F_2(X_2))=\frac{6}{\pi}\arcsin(\rho/2)$

Correlation Fallacies

Fallacy 1: F_1,F_2 and ρ uniquely determine F.
This is true for bivariate elliptical distributions, but wrong in general.
Example is as follows (**uncorrelated $\nRightarrow$ independent**)

- Consider the two risks $X_1=Z, X_2=ZV$ where V,Z are independent with $Z\sim\mathcal{N}(0,1)$ and $\mathbb{P}[V=-1]=\mathbb{P}[V=1]=1/2$.
  Then $X_1,X_2\sim\mathcal{N}(0,1)$ and $\rho(X_1,X_2)=\mathrm{Cov}(X_1,X_2)=\mathbb{E}[X_1X_2]\overset{\text{ind.}}{=}\mathbb{E}[V]\mathbb{E}\left[Z^2\right]=0$, but X_1 and X_2 are not independent. Indeed, $|X_1|=|X_2|=|Z|$ and so $\mathrm{Cov}(|X_1|,|X_2|)=\mathrm{Var}(|Z|)>0$.
  In particular, (X_1,X_2) is not bivariate normal.
- Consider $\left(X_1',X_2'\right)\sim\mathcal{N}_2(0,\mathbf{I}_2)$. Both $\left(X_1',X_2'\right)$ and (X_1,X_2) have $\mathcal{N}(0,1)$ margins and $\rho=0$, but the copula of $\left(X_1',X_2'\right)$ is Π and the copula of (X_1,X_2) is $C(u)=W(u)/2+M(u)/2$.
  V switches between perfectly positive and negative dependence!

Fallacy 2: For given marginal cdf's F_1,F_2, any $\rho\in[-1,1]$ is attainable.
This is true for elliptically distributed (X_1,X_2) with $\mathbb{E}\left[R^2\right]<\infty$ (as then $\mathrm{Corr}(X_1,X_2)=\mathbf{P}$), but wrong in general:

- If F_1 and F_2 are not of the same type (i.e., affine transformations of each other), then $\rho(X_1,X_2)=1$ is not attainable.
  Recall that $|\rho|=1$ **iff** there exist constants $a\in\mathbb{R}\backslash\{0\}, b\in\mathbb{R}$ such that $X_2=aX_1+b$ a.s.
- Hoeffding's identity

$$\begin{aligned}&\mathrm{Cov}(X_1,X_2)\\=&\iint_{\mathbb{R}^2}[F(x_1,x_2)-F_1(x_1)F_2(x_2)]dx_1dx_2\\=&\iint_{\mathbb{R}^2}[C(F_1(x_1),F_2(x_2))-F_1(x_1)F_2(x_2)]dx_1dx_2\end{aligned}$$

implies bounds on attainable ρ: $\rho\in[\rho_{\min},\rho_{\max}]$ ($\rho_{\min}$ is attained for $C=W$, $\rho_{\max}$ for $C=M$).

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 in this case, VaR_α is subadditive and $\rho=1$ implies that X_1,X_2 are comonotone. Moreover, VaR_α is always comonotone additive.
- Any super-additivity example $\mathrm{VaR}_\alpha(X_1+X_2)>\mathrm{VaR}_\alpha(X_1)+\mathrm{VaR}_\alpha(X_2)$ (the right-hand side is $\mathrm{VaR}_\alpha(X_1+X_2)$ under comonotonicity, which gives maximal correlation) serves as a counterexample.

Rank Correlation

Rank correlation coefficients are: (1) always defined; (2) invariant under strictly increasing transformations of the marginals (hence only depend on the underlying copula)

Def (Kendall's tau)

Let $X_j \sim F_j$ with continuous $F_j, j=1,2$, and (X_1', X_2') an independent copy of (X_1, X_2).
Kendall's tau is defined by

$$\begin{aligned}\rho_\tau :=& \mathbb{E}\left[\operatorname{sign}\left(\left(X_1-X_1'\right)\left(X_2-X_2'\right)\right)\right]\\ =&\mathbb{P}\left[\left(X_1-X_1'\right)\left(X_2-X_2'\right)>0\right]\\ &-\mathbb{P}\left[\left(X_1-X_1'\right)\left(X_2-X_2'\right)<0\right]\end{aligned}$$

where $\operatorname{sign}(x) = 1_{(0,\infty)}(x) - 1_{(-\infty,0)}(x)$.
By definition, Kendall's tau is the probability of concordance minus the probability of discordance.
An estimator of ρ_τ is provided by the sample version of Kendall's tau

$$r_\tau(n) = \frac{1}{\binom{n}{2}} \sum_{1\le i_1<i_2\le n} \operatorname{sign}\left(\left(X_{i_1,1}-X_{i_2,1}\right)\left(X_{i_1,2}-X_{i_2,2}\right)\right)$$

where $\left(X_{1,1}, X_{1,2}\right)\ldots,\left(X_{n,1},X_{n,2}\right)$ are n independent realizations of (X_1, X_2).

Proposition (Formula for Kendall's tau)

Assume (X_1, X_2) has copula C and continuous marginals F_1 and F_2.
Then $\rho_\tau = 4\iint_{[0,1]^2} C(u_1,u_2)\,dC(u_1,u_2) - 1 = 4\mathbb{E}C(U_1,U_2)-1$ where $(U_1,U_2)\sim C$.
In particular, ρ_τ only depends on the copula C of (X_1, X_2).

Proof: Let $\left(X_1', X_2'\right)$ be an independent copy of (X_1, X_2) and denote $U_j = F_j\left(X_j\right), U_j' = F_j\left(X_j'\right)$.
Then $U \sim U' \sim C$, and

$$\begin{aligned}&\rho_\tau\\ =&\mathbb{P}\left[\left(X_1-X_1'\right)\left(X_2-X_2'\right)>0\right]\\ &-\mathbb{P}\left[\left(X_1-X_1'\right)\left(X_2-X_2'\right)<0\right]\\ =&2\mathbb{P}\left[\left(X_1-X_1'\right)\left(X_2-X_2'\right)>0\right]-1\\ =&4\mathbb{P}\left[X_1<X_1', X_2<X_2'\right]-1\\ =&4\mathbb{P}\left[U_1<U_1', U_2<U_2'\right]-1\\ =&4\iint_{[0,1]^2}\mathbb{P}\left[U_1<u_1,U_2<u_2\right]dC(u_1,u_2)-1\\ =&4\iint_{[0,1]^2}C(u_1,u_2)\,dC(u_1,u_2)-1\end{aligned}$$

- For $C = \Pi$:

$$\rho_\tau = 4\iint_{[0,1]^2} u_1u_2\,du_1du_2 - 1 = 0$$

$\rightsquigarrow$ if X_1 and X_2 are independent, then $\rho_\tau = 0$

- For $C = M$:

$$\begin{aligned}\rho_\tau &= 4\iint_{[0,1]^2} M(u_1,u_2)\,dM(u_1,u_2) - 1\\ &= 4\mathbb{E}[U\wedge U] - 1 = 4\mathbb{E}[U] - 1 = 2-1 = 1\end{aligned}$$

$\rightsquigarrow$ the upper bound 1 is attained for any pair of continuous marginals F_1, F_2

- For $C = W$:

$$\begin{aligned}\rho_\tau &= 4\iint_{[0,1]^2} W(u_1,u_2)\,dW(u_1,u_2) - 1\\ &= 4\mathbb{E}\left[(U+(1-U)-1)^+\right] - 1 = -1\end{aligned}$$

$\rightsquigarrow$ the lower bound -1 is attained for any pair of continuous marginals F_1, F_2

Def (Spearman's rho)

Assume (X_1, X_2) has continuous marginals F_1 and F_2. Then Spearman's rho is defined by $\rho_S = \rho(F_1(X_1), F_2(X_2))$.
An estimator $r_S(n)$ is given by the sample correlation of $\left(\operatorname{rk}\left(X_{1,1}\right),\operatorname{rk}\left(X_{1,2}\right)\right),\ldots,\left(\operatorname{rk}\left(X_{n,1}\right),\operatorname{rk}\left(X_{n,2}\right)\right)$ where $\left(X_{1,1},X_{1,2}\right)\ldots,\left(X_{n,1},X_{n,2}\right)$ are n independent realizations of (X_1,X_2) and $\operatorname{rk}\left(X_{i,j}\right)$ is the rank of $X_{i,j}$ among $X_{1,j},\ldots,X_{n,j}$.

Proposition (Formula for Spearman's rho)

Assume (X_1, X_2) has copula C and continuous marginals F_1 and F_2.
Then $\rho_S = 12\iint_{[0,1]^2} C(u_1,u_2)\,du_1du_2 - 3 = 12\mathbb{E}C(U_1,U_2)-3$ where $(U_1,U_2)\sim\Pi$.
In particular, ρ_S only depends on the copula C of (X_1, X_2).
Proof:

$$\begin{aligned}\rho_S(X_1,X_2) &= \rho(F_1(X_1), F_2(X_2))\\ &= \frac{\operatorname{Cov}(F_1(X_1),F_2(X_2))}{\sqrt{\operatorname{Var}(F_1(X_1))\operatorname{Var}(F_2(X_2))}}\\ &= \frac{\iint_{[0,1]^2}(C(u_1,u_2)-u_1u_2)\,du_1du_2}{\operatorname{Var}(U)}\\ &= 12\iint_{[0,1]^2}(C(u_1,u_2)-u_1u_2)\,du_1du_2\\ &= 12\int_{[0,1]^2} C(u_1,u_2)\,du_1du_2 - 3\end{aligned}$$

- For $C = \Pi$: $\rho_\tau = \rho_S = 0 \rightsquigarrow$ if X_1 and X_2 are independent, then $\rho_\tau = \rho_S = 0$
- For $C = M$: $\rho_\tau = \rho_S = 1 \rightsquigarrow$ the upper bound 1 is attained for any pair of cont marginals F_1, F_2
- For $C = W$: $\rho_\tau = \rho_S = -1 \rightsquigarrow$ the lower bound -1 is attained for any pair of cont marginals F_1, F_2
- For $\kappa = \rho_\tau$ and $\kappa = \rho_S$, one has $\kappa = \pm 1$ if and only if X_1, X_2 are co-/counter-monotonic
- Fallacy 1 (F_1, F_2, ρ uniquely determine F) is not solved by replacing ρ with rank correlation coefficient κ (it is easy to construct different copulas with the same Kendall's tau, e.g. via Archimedean copulas)
- Fallacy 2 (for given continuous F_1, F_2, any $\rho \in [-1,1]$ is attainable) is solved. Set $F(x_1,x_2) = \lambda W(F_1(x_1),F_2(x_2)) + (1-\lambda)M(F_1(x_1),F_2(x_2))$. This is a model with $\rho_\tau = \rho_S = 1-2\lambda$ (choose $\lambda\in[0,1]$ as desired)
- Fallacy 3 (κ maximal $\Rightarrow \operatorname{VaR}_\alpha(X_1+X_2)$ maximal) is also not solved by rank correlation ($\kappa = 1$ corresponds to $C = M$, but this copula does not necessarily provide the largest $\operatorname{VaR}_\alpha(X_1+X_2)$; see the super-additivity examples)
- Also, in general, $\kappa = 0$ does not imply independence
- Nevertheless, rank correlations are useful to summarize dependence, to compare different dependence structures as well as for copula parameter calibration/estimation

Coefficients of Tail Dependence

Goal: Measure extremal dependence, that is, dependence in the joint tails.

Def (Coefficients of tail dependence)

Let (X_1, X_2) be a random vec with continuous marginals F_1 and F_2. Provided that the limits exist, the lower tail dependence coefficient λ_l and upper tail dependence coefficient λ_u of (X_1, X_2) are defined by $\lambda_l = \lim_{\alpha\downarrow 0}\mathbb{P}\left[X_2 \le q_{X_2}^-(\alpha)\mid X_1 \le q_{X_1}^-(\alpha)\right]$, $\lambda_u = \lim_{\alpha\uparrow 1}\mathbb{P}\left[X_2 > q_{X_2}^-(\alpha)\mid X_1 > q_{X_1}^-(\alpha)\right]$.
Note: as limits of (conditional) probabilities, λ_l and λ_u are in $[0,1]$

Def (Tail dependence and independence)

If $\lambda_l > 0\ (\lambda_u > 0)$, (X_1, X_2) is said to be lower (upper) tail dependent;
If $\lambda_l = 0\ (\lambda_u = 0)$, (X_1, X_2) is said to be lower (upper) tail independent.

- Tail dependence is a copula property, since

$$\begin{aligned}&\mathbb{P}\left[X_2\le q_{X_2}^-(\alpha)\mid X_1\le q_{X_1}^-(\alpha)\right]\\ =&\frac{\mathbb{P}\left[X_1\le q_{X_1}^-(\alpha), X_2\le q_{X_2}^-(\alpha)\right]}{\mathbb{P}\left[X_1\le q_{X_1}^-(\alpha)\right]}\\ =&\frac{F\left(q_{X_1}^-(\alpha), q_{X_2}^-(\alpha)\right)}{F_1\left(q_{X_1}^-(\alpha)\right)} = \frac{C(\alpha,\alpha)}{\alpha}, \alpha\in(0,1)\end{aligned}$$

- Lower tail dependence coeff: $\lambda_l = \lim_{\alpha\downarrow 0}\frac{C(\alpha,\alpha)}{\alpha}$.
- Upper tail dep coeff: $\lambda_u = 2 - \lim_{\alpha\uparrow 1}\frac{1-C(\alpha,\alpha)}{1-\alpha}$
- If $\alpha \mapsto C(\alpha,\alpha)$ is differentiable in a neighborhood of 0 and the limit exists, then $\lambda_l = \lim_{\alpha\downarrow 0}\frac{d}{d\alpha}C(\alpha,\alpha)$ (l'Hôpital's rule)
- If $(x,y)\mapsto C(x,y)$ is differentiable in a neighborhood of 0 and the limit exists, then $\lambda_l = \lim_{\alpha\downarrow 0}(\partial_1 C(\alpha,\alpha) + \partial_2 C(\alpha,\alpha))$ (chain rule)
- For all radially symmetric copulas (e.g. the bivariate C_P^{Ga} and $C_{\nu,P}^t$ copulas), $\lambda_l = \lambda_u =: \lambda$.
- For Archimedean copulas with strict ψ, $\lambda_l = 2\lim_{x\to\infty}\frac{\psi'(2x)}{\psi'(x)}$, $\lambda_u = 2 - 2\lim_{x\downarrow 0}\frac{\psi'(2x)}{\psi'(x)}$

$$\begin{aligned}\lambda_l &= \lim_{\alpha\downarrow 0}\frac{\psi\left(2\psi^{-1}(\alpha)\right)}{\alpha} = \lim_{x\to\infty}\frac{\psi(2x)}{\psi(x)}\\ &= 2\lim_{x\to\infty}\frac{\psi'(2x)}{\psi'(x)}\end{aligned}$$

$$\begin{aligned}\lambda_u &= 2 - \lim_{\alpha\uparrow 1}\frac{1-\psi\left(2\psi^{-1}(\alpha)\right)}{1-\alpha} = 2 - \lim_{x\downarrow 0}\frac{1-\psi(2x)}{1-\psi(x)}\\ &= 2 - 2\lim_{x\downarrow 0}\frac{\psi'(2x)}{\psi'(x)}\end{aligned}$$

- Clayton: $\lambda_l = 2^{-1/\theta}, \lambda_u = 0$
- Gumbel: $\lambda_l = 0, \lambda_u = 2 - 2^{1/\theta}$
- If $(x,y)\mapsto C(x,y)$ is differentiable and C is symmetric, then $\lambda_l = 2\lim_{\alpha\downarrow 0}\partial_1 C(\alpha,\alpha)$.
  Moreover,
  - $\partial_1 C(\alpha,\alpha) = \partial_1\int_0^\alpha\int_0^\alpha c(x,y)\,dxdy = \int_0^\alpha c(\alpha,y)\,dy$
  - $\int_0^1 c(\alpha,y)\,dy = \partial_1\int_0^\alpha\int_0^1 c(x,y)\,dxdy = \partial_1 C(\alpha,1) = \frac{d}{d\alpha}\alpha = 1$
  - $\int_0^\alpha c(\alpha,y)\,dy = \frac{\int_0^\alpha c(\alpha,y)dy}{\int_0^1 c(\alpha,y)dy} = \mathbb{P}[U_2\le\alpha\mid U_1=\alpha]$ if $(U_1,U_2)\sim C$
  - $\lambda_l = 2\lim_{\alpha\downarrow 0}\mathbb{P}[U_2\le\alpha\mid U_1=\alpha]$ for $(U_1,U_2)\sim C$
  - If $G(x) = \int_{-\infty}^x g(y)\,dy$ for a positive density g, then for $(X_1,X_2) = \left(G^{-1}(U_1), G^{-1}(U_2)\right)$, one has

$$\begin{aligned}\lambda_l &= 2\lim_{x\downarrow-\infty}\mathbb{P}[X_2\le x\mid X_1 = x]\\ &= 2\lim_{x\downarrow-\infty}\int_{-\infty}^x f_{X_2|X_1=x}(y)\,dy\end{aligned}$$

7.3 Normal mixture copulas

Tail Dependence

Coefficients of tail dependence

Let (X_1, X_2) be distributed according to a normal variance mixture and assume (w.l.o.g.) that $\boldsymbol{\mu} = (0,0)$ and $\mathbf{A}\mathbf{A}^\top = \Sigma = \mathbf{P} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. In this case, $F_1 = F_2$ and C is symmetric as well as radially symmetric. One thus obtains

$$\lambda = \lambda_l = \lambda_u = 2 \lim_{x \downarrow -\infty} \mathbb{P}[X_2 \le x \mid X_1 = x]$$
$$= 2 \lim_{x \downarrow -\infty} \int_{-\infty}^{x} f_{X_2|X_1=x}(y) dy$$

Example: tail dependence for the Gauss- and t-copula

- For $(X_1, X_2) \sim \mathcal{N}(0, \mathbf{P})$ for $\mathbf{P} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, one has $X_2 \mid X_1 = x \sim \mathcal{N}\left(\rho x, 1-\rho^2\right)$.
  Hence $\lambda = 2\lim_{x\downarrow-\infty} \mathbb{P}[X_2 \le x \mid X_1 = x] = 2\lim_{x\downarrow-\infty} \Phi\left(\frac{x(1-\rho)}{\sqrt{1-\rho^2}}\right) = 1_{\{\rho=1\}}$
  $\rightsquigarrow$ no tail dependence
- For $C^t_{\nu,P}$, one can show that $X_2 \mid X_1 = x \sim t_{\nu+1}\left(\rho x, \frac{(1-\rho^2)(\nu+x^2)}{\nu+1}\right)$, and thus
  $$\mathbb{P}[X_2 \le x \mid X_1 = x] = t_{\nu+1}\left(\frac{x(1-\rho)}{\sqrt{\frac{(1-\rho^2)(\nu+x^2)}{\nu+1}}}\right).$$
  So $\lambda = 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right)$
  $\rightsquigarrow$ tail dependence

Rank Correlations

Proposition (Spearman's rho for normal variance mixtures)

Let $X \sim M_2\left(0, P, \hat{F}_W\right)$ with $\mathbb{P}[X = 0] = 0, \rho = P_{12}$. Then $\rho_S = \frac{6}{\pi}\mathbb{E}\left[\arcsin\left(\frac{\rho W}{\sqrt{(W+\tilde{W})(W+\overline{W})}}\right)\right]$, for $W, \tilde{W}, \overline{W} \overset{\text{iid}}{\sim} F_W$ with Laplace-Stieltjes transform $\hat{F}_W$.

For Gauss copulas, $\rho_S = \frac{6}{\pi}\arcsin(\rho/2)$.

Proposition (Kendall's tau for elliptical dist)

Let $X \sim E_2(0, \mathbf{P}, \psi)$ with $\mathbb{P}[X = 0] = 0$ and $\rho = \mathbf{P}_{12}$. Then $\rho_\tau = \frac{2}{\pi}\arcsin\rho$.

Skewed Normal Mixture Copulas

- Skewed normal mixture copulas are the copulas of normal mixture distributions which are not elliptical, e.g. the skewed t-copula is the copula of a generalized hyperbolic distribution
- It can be sampled as other implicit copulas; (the transformations of the margins requires numerical integration of a skewed t-density)
- The main advantage of such a copula over $C^t_{\nu,P}$ is its radial asymmetry (e.g. for modeling $\lambda_l \ne \lambda_u$)

Grouped Normal Mixture Copulas

A **grouped normal mixture copula** is the copula of a random vector of the form $X = \left(\sqrt{W_1}Y_1, \dots, \sqrt{W_1}Y_{k_1}, \dots, \sqrt{W_n}Y_{k_{n-1}+1}, \dots, \sqrt{W_n}Y_{k_n}\right)$ where (1) $\mathbf{Y} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{P})$ for a correlation matrix $\mathbf{P}$; i.e. $\mathbf{Y} \overset{(d)}{=} \mathbf{A}\mathbf{Z}$ for $\mathbf{A}\mathbf{A}^\top = \mathbf{P}$; (2) $1 \le k_1 \le k_2 \le \ldots \le k_n = d$; (3) $W_1, \dots, W_n$ are non-negative comonotone RVs.

Example:

$$\mathbf{X} = \left(\sqrt{W_1}Y_1, \dots, \sqrt{W_1}Y_{k_1}, \dots, \sqrt{W_n}Y_{k_{n-1}+1}, \dots, \sqrt{W_n}Y_{k_n}\right)$$

where

- $\mathbf{Y} \sim \mathcal{N}_d(0, \mathbf{P})$ for a correlation matrix $\mathbf{P}$. i.e. $\mathbf{Y} \overset{(d)}{=} \mathbf{A}\mathbf{Z}$ for $\mathbf{A}\mathbf{A}^\top = \mathbf{P}$
- $1 \le k_1 \le k_2 \le \dots \le k_n = d$
- $W_1, \dots, W_n$ are comonotone such that $W_j = 1/G_j$ for $G_j \sim \Gamma\left(\nu_j/2, \nu_j/2\right), j = 1, \dots, n$

The marginals are t_{ν_j}-distributed, $j = 1, \dots, n$.

$$\mathbf{U} = \left(t_{\nu_1}(X_1), \dots, t_{\nu_1}\left(X_{k_1}\right), \dots, t_{\nu_n}\left(X_{k_{n-1}+1}\right), \dots, t_{\nu_n}\left(X_{k_n}\right)\right)$$

follows a **grouped t-copula**.
For $k_n = d$, **grouped t-copulas** are also known as **generalized t-copulas**.

7.4 Archimedean copulas

Recall that an **Archimedean generator** ψ is a function $\psi : [0, \infty) \to [0, 1]$ satisfying

- $\psi(0) = 1$
- $\lim_{x\to\infty} \psi(x) = 0$
- ψ is continuous, non-increasing and strictly decreasing on $[0, \inf\{x : \psi(x) = 0\}]$

The set of all generators is denoted by Ψ.

Bivariate Archimedean Copulas

For $\psi \in \Psi, C(u_1, u_2) = \psi\left(\psi^{-1}(u_1) + \psi^{-1}(u_2)\right)$ is a copula **iff** ψ is convex.

- For a strict and twice-continuously differentiable ψ, one can show that
  $$\rho_\tau = 1 - 4\int_0^\infty x\left(\psi'(x)\right)^2 dx = 1 + 4\int_0^1 \frac{\psi^{-1}(x)}{\left(\psi^{-1}(x)\right)'} dx$$
- If ψ is strict, $\lambda_l = 2\lim_{x\to\infty} \frac{\psi'(2x)}{\psi'(x)}$ and $\lambda_u = 2 - 2\lim_{x\downarrow 0} \frac{\psi'(2x)}{\psi'(x)}$

Multivariate Archimedean Copulas

ψ is **completely monotone (c.m.)** if $(-1)^k \psi^{(k)}(x) \ge 0$ for all $x \in (0, \infty)$ and all $k \in \mathbb{N}_0$. The set of all c.m. generators is denoted by Ψ_∞.

Thm (Kimberling)

For $\psi \in \Psi, C(\mathbf{u}) = \psi\left(\sum_{j=1}^d \psi^{-1}\left(u_j\right)\right)$ is a copula for all $d \ge 2$ **iff** $\psi \in \Psi_\infty$.

Thm (Bernstein)

A function $\psi : [0, \infty) \to [0, 1]$ is completely monotone **iff** $\psi(x) = \mathbb{E}[\exp(-xV)]$ for a non-negative random variable $V \sim G$ with $G(0) = 0$

Notation: $\psi = \hat{G}$ (**Laplace transform**)

It can be shown that to generate a d-dimensional copula, it is enough for ψ to be d-monotone.

Proposition (Stochastic representation)

Let $\psi \in \Psi_\infty$ such that $\psi = \hat{G}$. Let $V \sim G$ and $E_1, \dots, E_d \overset{\text{iid}}{\sim} \text{Exp}(1)$ ind of V. Then

(1) The survival copula $\hat{C}$ of $\mathbf{X} = (E_1/V, \dots, E_d/V)$ is Archimedean with generator ψ

(2) $\mathbf{U} = (\psi(X_1), \dots, \psi(X_d)) \sim \hat{C}$ and the U_j 's are conditionally independent given V with
$$\mathbb{P}\left[U_j \le u \mid V = v\right] = \exp\left(-v\psi^{-1}(u)\right)$$

Proof:

(1) The joint survival function of $\mathbf{X}$ is given by
$$\begin{aligned}
\overline{F}(x) &= \mathbb{P}\left[X_j > x_j \text{ for all } j\right] \\
&= \int_0^\infty \mathbb{P}\left[E_j/V > x_j \text{ for all } j \mid V = v\right] dG(v) \\
&= \int_0^\infty \mathbb{P}\left[E_j > vx_j \text{ for all } j\right] dG(v) \\
&= \int_0^\infty \prod_{j=1}^d \exp\left(-vx_j\right) dG(v) \\
&= \int_0^\infty \exp\left(-v\sum_{j=1}^d x_j\right) dG(v) = \psi\left(\sum_{j=1}^d x_j\right)
\end{aligned}$$
$$\overline{F}_j\left(x_j\right) = \overline{F}\left(0, \dots, 0, x_j, 0, \dots, 0\right) = \psi\left(x_j\right)$$
$$\begin{aligned}
\hat{C}(u) &= \overline{F}\left(\overline{F}_1^{-1}(u_1), \dots, \overline{F}_d^{-1}(u_d)\right) \\
&= \psi\left(\sum_{j=1}^d \psi^{-1}\left(u_j\right)\right)
\end{aligned}$$

(2) $\mathbb{P}[U \le u] = \mathbb{P}\left[\psi\left(X_j\right) \le u_j \text{ for all } j\right] = \mathbb{P}\left[X_j > \psi^{-1}\left(u_j\right) \text{ for all } j\right] = \psi\left(\sum_{j=1}^d \psi^{-1}\left(u_j\right)\right)$
Conditional independence is clear by construction, and
$$\begin{aligned}
\mathbb{P}\left[U_j \le u \mid V = v\right] &= \mathbb{P}\left[\psi\left(X_j\right) \le u \mid V = v\right] \\
&= \mathbb{P}\left[X_j > \psi^{-1}(u) \mid V = v\right] = \mathbb{P}\left[E_j > v\psi^{-1}(u)\right] \\
&= \exp\left(-v\psi^{-1}(u)\right)
\end{aligned}$$

Algorithm (Marshall and Olkin)

(1) Sample $V \sim G$, where $\hat{G} = \psi$

(2) Sample $E_1, \dots, E_d \overset{\text{iid}}{\sim} \text{Exp}(1)$ ind of V

(3) Return $U = (\psi(E_1/V), \dots, \psi(E_d/V))$

7.5 Fitting copulas to data

- Let $X, X_1, \dots, X_n$ be independent d-dimensional random vectors with cdf F, continuous margins $F_1, \dots, F_d$ and copula C
- Let $x_1, \dots, x_n \in \mathbb{R}^d$ be realizations of $X_1, \dots, X_n$
- Assume (1) $F_j = F_j\left(\cdot; \theta_j\right)$ for some $\theta_j \in \Theta_j, j = 1, \dots, d$; (2) $C = C(\cdot; \theta_C)$ for some $\theta_C \in \Theta_C$
  The true but unknown parameter vector $\theta^* = \left(\theta^*_C, \theta^*_1, \dots, \theta^*_d\right)$ has to be estimated
- Here, we focus particularly on θ_C. Whenever necessary, we assume that the margins $F_1, \dots, F_d$ and the copula C are absolutely continuous with corresponding densities $f_1, \dots, f_d$ and c, respectively

Method-of-Moments Using Rank Correlation

- We focus on one-parameter copulas here
- For $d = 2$, Genest and Rivest suggested estimating θ_C by choosing it so that $\rho_\tau(\theta_C) = r_\tau(n)$, that is,
  $$\hat{\theta}^{\text{IKTE}}_{n,C} = \rho_\tau^{-1}(r_\tau(n))$$
  (**inversion of Kendall's tau estimator (IKTE)**)
  where $\rho_\tau(\cdot)$ denotes Kendall's tau as a function of θ and $r_\tau(n)$ is the sample version of Kendall's tau
- The standardized dispersion matrix P for **elliptical copulas** can be estimated via pairwise inversion of Kendall's tau. If $r_\tau^{j_1 j_2}(n)$ denotes the sample version of Kendall's tau for data pair (j_1, j_2), then $\hat{p}^{\text{IKTE}}_{n,j_1,j_2} = \sin\left(\pi r_\tau^{j_1,j_2}(n)/2\right)$.
  Recall: $\rho_\tau = \frac{2}{\pi}\arcsin\rho$ for elliptical distr.
  (A correction might be needed for obtaining a proper correlation matrix $\mathbf{P}$; that is, one that is positive semi-definite)
- For Gauss copulas, it is preferable to use Spearman's rho based on
  $$\rho_S = \frac{6}{\pi}\arcsin\frac{\rho}{2} \approx \rho$$
  The latter approximation error is relatively small, so that the matrix of pairwise sample versions of Spearman's rho is an estimator for $\mathbf{P}$
- For t-copulas, $\hat{P}^{\text{IKTE}}_n$ can be used to estimate $\mathbf{P}$ and then ν can be estimated via its MLE based on $\hat{P}^{\text{IKTE}}_n$.

Forming a Pseudo-Sample from the Copula

- $X_1,\dots,X_n$ typically does not have $U(0,1)$ margins. For applying the "copula approach", one needs **pseudo-observations** from C
- For instance, $\hat{\mathbf{U}}_i = (\hat{U}_{i,1},\dots,\hat{U}_{i,d}) = (\hat{F}_1(X_{i,1}),\dots,\hat{F}_d(X_{i,d})), i = 1,\dots,n$, where $\hat{F}_j$ is an estimator of F_j
- Note that $\hat{U}_1,\dots,\hat{U}_n$ are typically neither independent (even if $X_1,\dots,X_n$ are) nor perfectly $U(0,1)$

Different ways of estimating $\hat{F}_j$

(1) **Non-parametric estimators** with scaled empirical df's.

The empirical cf$\hat{F}_j$ is given by

$$\hat{F}_j(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_{i,j}\le x\}}$$

Set $\hat{U}_{i,j} = \frac{n}{n+1}\hat{F}_j(X_{i,j}) = \frac{R_{i,j}}{n+1}$ **(1)**, where $R_{i,j}$ denotes the rank of $X_{i,j}$ among all $X_{1,j},\dots,X_{n,j}$

(the rescaling $n/(n+1)$ is used to avoid density evaluation on the boundary of $[0,1]^d$)

(2) **Parametric estimators** (such as Student- t, Pareto, etc.; typically if n is small).

In this case, one often still uses **(1)** for estimating θ_C

(to keep the error due to misspecification of the margins small)

(3) **Semi-parametric estimators** (e.g. EVT-based: bodies are modeled empirically, tails semi-parametrically via GPD)

Maximum Likelihood Estimation

The Classical Maximum Likelihood Estimator

- By Sklar's Theorem, the density of F is given by

$$\begin{aligned} &f(x;\theta)\\ =&c(F_1(x_1;\theta_1),\dots,F_d(x_d;\theta_d);\theta_C)\prod_{j=1}^{d} f_j(x_j;\theta_j) \end{aligned}$$

- The log-likelihood based on $X_1,\dots,X_n$ is thus

$$\begin{aligned} \ell(X_1,\dots,X_n;\theta) &= \sum_{i=1}^{n}\ell(X_i;\theta)\\ &= \sum_{i=1}^{n}\ell_C(F_1(X_{i,1};\theta_1),\dots,F_d(X_{i,d};\theta_d);\theta_C) \end{aligned}$$

where $\ell_C(u_1,\dots,u_d;\theta_C) = \log c(u_1,\dots,u_d;\theta_C)$ and $\ell_j(x;\theta_j) = \log f_j(x;\theta_j), j = 1,\dots,d$

- The maximum likelihood estimator (MLE) of θ is

$$\hat{\theta}_n^{\text{MLE}} = \underset{\theta\in\Theta}{\operatorname{argmax}}\,\ell(X_1,\dots,X_n;\theta).$$

This optimization is typically done by numerical means. But it can be quite demanding, especially in high dimensions.

The Inference Functions for Margins Estimator (IFME)

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

(1) For $j = 1,\dots,d$, estimate θ_j by its MLE $\hat{\theta}_{n,j}^{\text{MLE}}$;

(2) Estimate θ_C by

$$\hat{\theta}_{n,C}^{\text{IFME}} = \underset{\theta_C\in\Theta_C}{\operatorname{argmax}}\,\ell(X_1,\dots,X_n;\hat{\theta}_{n,1}^{\text{MLE}},\dots,\hat{\theta}_{n,d}^{\text{MLE}},\theta_C)$$

The **inference functions for margins estimator (IFME)** of θ is thus

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

- This is typically much easier to compute than $\hat{\theta}_n^{\text{MLE}}$ while providing good results
- $\hat{\theta}_n^{\text{IFME}}$ can also be used as initial value for a numerical evaluation of $\hat{\theta}_n^{\text{MLE}}$

Example: Suppose $X_j \sim N(\mu_j,\sigma_j^2), j = 1,\dots,d$ for $d = 100$, and C has one parameter

- MLE requires to solve a 201-dimensional optimization problem
- 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 df's to estimate θ_C (see MPLE below) in order to avoid misspecification of the margins (if n is sufficiently large)

The Maximum Pseudo-Likelihood Estimator (MPLE)

- The maximum pseudo-likelihood estimator (MPLE), introduced by Genest et al. (1995), works similarly to θ_n^{IFME}, but estimates the margins non-parametrically:

(1) Compute rank-based pseudo-observations $\hat{U}_{i,j} = R_{i,j}/(n+1)$;

(2) Estimate θ_C by

$$\begin{aligned} \theta_{n,C}^{\text{MPLE}} &= \underset{\theta_C\in\Theta_C}{\operatorname{argmax}}\sum_{i=1}^{n}\ell_C(\hat{U}_{i,1},\dots,\hat{U}_{i,d};\theta_C)\\ &= \underset{\theta_C\in\Theta_C}{\operatorname{argmax}}\sum_{i=1}^{n}\log c(\hat{U}_i;\theta_C) \end{aligned}$$

- Genest and Werker (2002) show that $\hat{\theta}_{n,C}^{\text{MPLE}}$ is not asymptotically efficient in general
- Kim et al. (2007) compare $\hat{\theta}_n^{\text{MLE}},\hat{\theta}_n^{\text{IFME}}$ and $\hat{\theta}_{n,C}^{\text{MPLE}}$ in a simulation study ($d = 2$ only!) and argue in favor of $\hat{\theta}_{n,C}^{\text{MPLE}}$ overall, especially w.r.t. robustness to misspecification of the margins

Example: fitting the Gauss copula

- The log-likelihood ℓ_C is

$$\ell_C(\hat{U}_1,\dots,\hat{U}_n;P) = \sum_{i=1}^{n}\ell_C(\hat{U}_i;P) = \sum_{i=1}^{n}\log c_P^{Ga}(\hat{U}_i)$$

For maximization over all correlation matrices P, one can use the Cholesky factor A as reparametrization and maximize over all lower triangular matrices A with 1 s on the diagonal this is still $\mathcal{O}(d^2)$

- Alternatively, one can use pairwise inversion of Spearman's rho or Kendall's tau

Example: fitting the t-copula

- For small d, maximize the likelihood over all correlation matrices (as in the Gauss copula case) and the degree of freedom ν
- For moderate/larger d, (1) Estimate P via pairwise inversion of Kendall's tau (2) Plug $\hat{P}$ into the likelihood and maximize it w.r.t. ν to obtain $\hat{\nu}_n$

Example: correlation estimation for heavy-tailed data

Consider $n = 3000$ realizations of independent samples of size 90 from $t_2\left(3,0,\begin{pmatrix}1 & 0.5\\ 0.5 & 1\end{pmatrix}\right) \rightsquigarrow$ linear correlation $\rho = 0.5$

Shall we estimate ρ via **sample correlation** or via **inversion of Kendall's tau**? The variance of the latter is smaller!

Estimation, Goodness-of-fit & Model Selection

Estimation is only one side of the coin. The other is **goodness-of-fit** (i.e. to find out whether the estimated model indeed represents the given data well) and **model selection** (i.e. to decide which model is best among different adequate fitted models). **Goodness-of-fit** can be (computationally) challenging, particularly for large d.

8 Operational Risk

Definition of Operational Risk

The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events.

This definition includes

- **Legal Risk** – financial loss that can result from lack of awareness or misunderstanding of, ambiguity in, or reckless indifference to, the way law and regulation apply to your business, its relationships, processes, products and services

but excludes

- **Strategic Risk** – loss arising from a poor strategic business decision
- **Reputational Risk** – damage to an organization through loss of its reputation or standing

Basel Op Risk categories

(1) **Internal Fraud** – misappropriation of assets, tax evasion, intentional mismarking of positions, bribery

(2) **External Fraud** – theft of information, hacking damage, third-party theft and forgery

(3) **Employment Practices and Workplace Safety** – discrimination, workers compensation

(4) **Clients, Products, and Business Practice** – market manipulation, product defects

(5) **Damage to Physical Assets** – natural disasters, terrorism, vandalism

(6) **Business Disruption and Systems Failures** – utility disruptions, software failures, hardware failures

(7) **Execution, Delivery, and Process Management** – data entry errors, accounting errors

8.1 Basel Pillar 1 - Minimal Capital Requirements

Regulatory Capital (RC)

Banks must hold a regulatory minimum capital to absorb losses from Op Risk. Currently, there are **three approaches**.

Basic Indicator Approach (BIA)

The Regulatory Capital under the BIA equals 15% of the average annual gross income over the previous three years where it was positive, i.e.,

$$\text{RC}_t^{\text{BIA}} = 15\%\cdot\left(\sum_{k=1}^{3}\max\{\text{GI}_{t-k},0\}\right)\Big/\left(\sum_{k=1}^{3}\mathbb{I}_{\text{GI}_{t-k}>0}\right)$$

where GI_{t-k} is the gross income in year $t-k$.

- For regional, non-complex firms
- Not risk-sensitive

The Standardized Approach (TSA)

The TSA is like the BIA, but the calculation is performed separately for each business line with different weights, i.e.,

$$\text{RC}_t^{\text{TSA}} = \frac{1}{3}\cdot\sum_{k=1}^{3}\max\left\{\sum_{b=1}^{8}\beta_b\text{GI}_{t-k}^{b},0\right\}$$

where $\text{GI}^b t-k$ is the gross income in year $t-k$ of business line b and β_b is its weight. The 8 business lines and their weights are (note the sum of weights is equal $1.2 = 8\times 15\%$):

Corporate finance 18%	Payment & Settlemer
Trading & Sales 18%	Agency Services 15%
Retail banking 12%	Asset management 1
Commercial banking 15%	Retail brokerage 12%

- Expected for most financial services firms
- Not risk-sensitive

Advanced Measurement Approach (AMA)

The Regulatory Capital is equal to the **Op Risk loss that is exceeded only once in 1000 years**, i.e., $\mathrm{VaR}_{0.999}(L)$, where the random variable L is the annual Op Risk loss.
Common approach to model L, taken in large banks, is the Loss Distribution Approach (LDA)
Allows banks to use their internally generated risk estimates, based on extensive Supervisory Guidance:
(1) Operational Risk - Supervisory Guidelines for the Advanced Measurement Approaches, Basel Committee on Banking Supervision
(2) Supervisory Guidance for Data, Modeling, and Model Risk Management Under the Operational Risk Advanced Measurement Approaches, FED ... (and more)

- Is currently in place at many of the global and systemically important banks (e.g., UBS)
- Requires prior regulatory approval
- Involves complex statistical modelling, allows for flexibility

Loss Distribution Approach (LDA) within the AMA Framework

Define Units of Measure (UoM)

A UoM typically combines business lines and loss event types, e.g., Investment Bank and Fraud

Model the annual UoM loss as compound sum of Frequency and Severity

For each $U \circ Mu, u \in \{1, 2, \ldots, U\}$, annual loss $L_u = \sum_{k=1}^{N_u} X_{k,u}$, where $\left\{X_{k,u} : k = 1, 2, \ldots, N_u\right\}$ are i.i.d and independent from N_u, N_u is the number of losses in UoM u per year (Frequency) and $X_{k,u}$ is the amount of the k-th loss in UoM u (Severity)

Aggregate the annual UoM losses into an annual loss distribution

The annual Op Risk loss L is then $L = \sum_{u=1}^{U} L_u$, where dependence structure of $(L_1, L_2, \ldots, L_U)$ is given by Copula C

Very challenging modelling problem

To estimate/justify: Segmentation into UoM / Frequency Distribution / Severity Distribution / Copula

The Four Data Elements for LDA Estimation

1. ILD; 2. ELD; 3. SA; 4. BEICF
More objective but backward-looking:
1. **Internal Operational Loss Event Data (ILD)**: Most relevant to the banks particular case, and well known
2. **External Operational Loss Event Data (ELD)**:

- Data consortia, e.g., Operational Riskdata eXchange Association (homogeneous classification standards, data relevance)
- Publicly available data, e.g., media or annual reports (reporting bias)

Forward-looking but more subjective:
3. **Scenario Analysis (SA)**:

- Systematic process of obtaining expert opinions on the likelihood and loss impact of plausible, high-severity operational losses, typically developed through workshops
- Expert biases (overconfidence, anchoring, . . .) and subjectivity

3. **Business Environment and Internal Control Factors (BEICF)**:

- Indicators designed to provide a forward-looking assessment of a banking organization's business risk factors and internal control environment (impact of discontinuing a line of business, a change in the internal control environment, . . .)
- Might be used to adjust operational risk exposure

Modelling Options within the LDA Framework

- Frequency: Poisson, Negative Binomial
- Severity: Log-Normal, Log-Gamma, Generalized Pareto, . . .
- Copula / Dependence
  - Dependence between annual losses, $L_u, u \in \{1, 2, \ldots, U\}$, vs dependence on frequency / severity level
  - Copulas: t, Clayton, Gumbel, Frank, ...
- Use of the four data elements (ILD, ELD, SA, BEICF): – Filtering ELD to remove non-relevant events; – Scaling ELD to account for differences in size or business activities; – Mixing data vs mixing distributions, e.g., fit distribution to ILD plus weighted ELD vs fit distributions to both ILD and ELD and mix the densities; – Benchmarking, e.g., compare ILD based main model with ELD based challenger model – Build an own SA distribution vs SA based adjustments; – Bayesian approach: use SA distributions as prior and calculate posterior given ILD and ELD; – Parameter adjustments based on BEICF

Criticism of the AMA

- Comparability of AMA minimum capital figures is questionable due to the full methodological freedom within the LDA
- How reliable are the quantitative estimates? 1-in-1000-year loss vs twenty years of ILD!
  - Limited availability of data / high confidence levels → uncertainty / instability in estimates
  - Over-fitting and extrapolation challenges

Standardized Measurement Approach (SMA)

- A simpler and more comparable approach will be implemented effective 1 January 2023*, see Section Minimum capital requirements for operational risk in Basel III: Finalising post-crisis reforms from the Basel Committee on Banking Supervision
- The SMA combines the Business Indicator, a simple financial statement proxy of operational risk exposure, with bankspecific Internal Loss Multiplier (based on internal operational loss data) to provide some incentive for banks to improve their operational risk management
- Nevertheless, comparability of capital charges remains a concern because
  - the collection of operational loss data is still determined by individual bank rules (no common standard),
  - national regulators may grant exclusion of (parts of) the loss history from the calculation (due to backward-looking nature of the SMA).

8.2 Basel Pillar 2 - Supervisory Review of Capital Adequacy

Supervisory Review of Capital Adequacy

- In the Internal Capital Adequacy Assessment Process, a bank needs to assess (among other things) whether it considers its capital adequate to cover the level and nature of the risks to which it is exposed
- This includes the risk types from Pillar 1 (Credit Risk, Market Risk and Op Risk), but extends to every possible risk type and their aggregation / diversification

Stress Testing

Stress Testing: analysis to determine whether there is enough capital to withstand the impact of an unfavorable economic scenario, including the causality chain by which losses would arise if the scenario were to unfold

Modelling Op Risk Stress Loss

1. Leverage the Loss Distribution Approach by choosing a particular quantile to represent the loss under stress

- Choice of the quantile is difficult to substantiate, which has led all CCAR banks to move away from this approach.

2. Regressions and case studies / scenarios analysis

- Quarterly operational losses may be linked to macroeconomic risk factor in certain cases, e.g., losses in "Execution, delivery, and process management" or "External fraud" to transaction volumes, real estate/house prices, GDP, and delinquency or unemployment rates.
- A possible modelling approach is to use generalized linear models with compound distributions (e.g., Poisson-Gamma) and log-link functions. The stability of the model and the lack of sufficient predictive power can be a challenge.
- A pool of candidate models is usually identified and the final model is selected by subject-matter experts and business intuition. If no economically meaningful model is found, fallback estimates may be used, e.g., historical average losses.
- To capture the possibility of rare (or not yet experienced) events, to include risk controls and mitigation efforts, to include a more "forward-looking" view, to evaluate the vulnerabilities of the bank identified during a risk identification exercise and to overcome some of the challenges when building a quantitative model, case study/scenario analysis is conducted by experts in workshops and the corresponding loss estimates (or loss estimate refinements) are included in the projections.

Economic Capital Calculation

- Economic Capital: amount of capital required to ensure solvency over a year with a pre-specified probability (e.g., 95% or 99.9%).
- A common approach to generate the annual loss distribution is to model risk types via their marginal loss distributions and then to aggregate them using copulas, see picture. For the Op Risk marginal distribution, the LDA can be used again (if there is already AMA LDA model).
- Another approach is to simulate many economic scenarios (risk drivers) consistently, and then to use methods like the ones used in Stress Testing to calculate the annual loss for each scenario.

8.3 Monitoring & Surveillance of Op Risks - Compliance Models

Compliance Models – Landscape

- **Adverse Media Screening**: Identify financial crime relevant news articles from various media sources concerning UBS clients or prospects
- **Sanctions Screening**: Identify references to sanctioned entities, individuals or regions within payment messages in order to prevent the transaction
- **AML Client Risk Rating**: Produce AML risk ratings which drive the frequency/depth of periodic client reviews and the level of alerting thresholds in downstream transaction monitoring systems
- **AML Transaction Monitoring**: Detect suspicious client transactions relating to money laundering (e.g., changes in transaction behavior, transactions involving high risk jurisdictions, flow through)
- **Communication Monitoring**: Detect potential compliance breaches in electronic messages (chats, e-mails, etc.) and audio communication (landline, mobile, Skype) of targeted UBS employees
- **Trade Surveillance**: Detect potential cases of market misconduct (e.g., insider trading, front running and trades away from the market price)

Compliance Models – Characteristics and Testing

- Use
  - Many models monitor key operational risks (Financial Crime, Market Conduct)
  - Alerts go through an expert review process and might ultimately lead to a regulatory filing
- Input data
  - Large amounts of data (trades, orders, text, audio, transactions, payments, client data), typically sourced from core systems
  - Processing is usually automated
- Methodology
  - Monthly, daily or event based execution of the alerting logic
  - Many submodels based on rules with many tunable parameters, statistical anomaly detection and/or Machine Learning
- Implementation
  - Inhouse built systems as well as on- and off-premise vendor solutions

- Implementation under resposibility of the IT department

- **Key model risk are false negatives (Type II error)**: The model does not produce an alert when it should have (false alerts "only" lead to extra effort and are well controlled through alert review)
- **Key testing / controls**:
  - Regular reviews of non-alerting cases / Below-the-Line testing
  - Regular testing with synthetic data
  - Regular coverage assessments (regulatory/internal requirements)

9 Exercises

Ex1 Multi t-dist

(1) Does there exist a two-dimensional random vector with a $t_2(\nu,\mu,\Sigma)$-distribution such that Σ is invertible and the components are independent of each other?

Solution: No. If $X \sim t_2(\nu,\mu,\Sigma)$ for an invertible 2×2 matrix Σ, it has a two-dimensional density of the form

$$f_X(x) = c\left(1+\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{\nu}\right)^{-\frac{\nu+2}{2}}, x\in\mathbb{R}^2$$

for a normalizing constant $c>0$. On the other hand, $X_1 \sim t_1(\nu,\mu_1,\Sigma_{11})$ and $X_2 \sim t_1(\nu,\mu_2,\Sigma_{22})$. In particular, they have one-dimensional densities

$$f_1(x_1) = c_1\left(1+\frac{(x_1-\mu_1)^2}{\nu\Sigma_{11}}\right)^{-\frac{\nu+1}{2}}$$

$$f_2(x_2) = c_2\left(1+\frac{(x_2-\mu_2)^2}{\nu\Sigma_{22}}\right)^{-\frac{\nu+1}{2}}$$

for normalizing contants $c_1,c_2>0$. So even if Σ is diagonal, $f_X(x)$ is not of the form $f_1(x_1)f_2(x_2)$, which shows that X_1 and X_2 are not independent.

(2) Does there exist a two-dimensional random vector with a $t_2(\nu,\mu,\Sigma)$-distribution such that the components are independent of each other?

Solution: Yes. But this is only possible if Σ is not invertible. The simplest case is $X=(0,0)\sim t_2(\nu,0,0)$

A less degenerate (but still degenerate) case is $X=(X_1,X_2)$, where $X_1=\mu_1+\sqrt{W}Z_1$ and $X_2=\mu_2$ for a deterministic vector $\mu=(\mu_1,\mu_2)\in\mathbb{R}^2$ and independent random variables $Z_1\sim N(0,1)$ and $W=1/G$ for $G\sim\Gamma(\nu/2,\nu/2)$. Then X_1 and X_2 are independent, and $X=(X_1,X_2)$ can be written as

$$X=\mu+\sqrt{W}AZ \quad \text{for} \quad A=\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$$

and a two-dimensional standard normal random vector $Z=(Z_1,Z_2)$ independent of W. So, $X\sim t_2(\nu,\mu,\Sigma)$ for $\Sigma=AA^T=A$.

(3) Does there exist a two-dimensional random vector such that both components have a standard one-dimensional t-distribution and are independent of each other?

Solution: Yes. Denote by f_ν the density of the standard one-dimensional t-distribution with $\nu>0$ degrees of freedom. For all $\nu_1,\nu_2>0$, $f_{\nu_1,\nu_2}(x_1,x_2)=f_{\nu_1}(x_1)f_{\nu_2}(x_2)$ is the density of a two-dimensional random vector whose components are independent and have standard one-dimensional t-distributions.

(4) Does there exist a $t_2(\nu,\mu,\Sigma)$-distribution that is spherical? Solution: Yes. The simplest example is again $t_2(\nu,0,0)$. More generally, $t_2\left(\nu,0,\sigma^2I_2\right)$ is spherical for all $\nu>0$ and $\sigma\geq 0$. Indeed, it has a representation of the form $X=\sqrt{W}\sigma Z$ for $Z\sim N_2(0,I_2)$ and an independent $W=1/G$ such that $G\sim\Gamma(\nu/2,\nu/2)$. Since Z is spherical, one has for every orthogonal 2×2 matrix U,

$$UX=\sqrt{W}\sigma UZ \overset{(d)}{=} \sqrt{W}\sigma Z = X,$$

which shows that $t_2\left(\nu,0,\sigma^2I_2\right)$ is spherical.

Ex2

Give a two-dimensional elliptical distribution that is not a normal or a t-distribution.

Solution: Let $Z\sim N_2(0,I_2)$ and W an independent non-negative random variable.| Then the distribution of $X=\sqrt{W}Z$ is spherical and, as a consequence, also elliptical. If W is constant, then X is normal, and if $W=1/G$ for $G\sim\Gamma(\nu/2,\nu/2)$, X has a 2-dimensional t distribution with ν degrees of freedom. In all other cases, X is neither normal nor t-distributed.
E.g. if W takes $k\geq 2$ different values with positive probabilities, X has a k point normal variance mixture distribution, which is neither normal nor a t-distribution.

Ex3 Fréchet-Hoeffding Bounds

a) Let X be an $\operatorname{Exp}(\lambda)$-distributed random variable for a parameter $\lambda>0$. Calculate the distribution function and the moments of $Y=\exp(X)$.
b) Y have a density? If yes, can you compute it?
c) Now, consider a two-dimensional random vector (X_1,X_2) such that $X_i\sim\operatorname{Exp}(\lambda_i)$ for parameters $\lambda_i>0, i=1,2$. Under which conditions does the linear correlation between $Y_1=\exp(X_1)$ and $Y_2=\exp(X_2)$ exist?
d) Assume $\lambda_1=3$ and $\lambda_2=4$. What is the range of possible correlations between Y_1 and Y_2?

Solution: a) $F_Y(y)=\mathbb{P}[\exp(X)\leq y]=F_X(\log(y))=1-y^{-\lambda}$ for all $y>1$ and $F_Y(y)=0$ for all $y\leq 1$.
$\mathbb{E}\left[Y^k\right]=\lambda\int_0^\infty e^{kz}e^{-\lambda z}dz=\frac{\lambda}{k-\lambda}e^{(k-\lambda)z}\Big|_0^\infty=\frac{\lambda}{\lambda-k}$ for all $k\in\mathbb{N}$ such that $k<\lambda$ and otherwise $\mathbb{E}\left[Y^k\right]=\infty$.
b) Since the cdf F_Y from a) is smooth on $(1,\infty)$ its pdf is given by $f_Y(y)=\frac{dF_Y}{dy}(y)=\frac{\lambda}{y^{\lambda+1}}$ for all $y>1$ and otherwise vanishes.
c) The linear correlation of X_1,X_2 exists if $X_i\in L^2(\mathbb{P})$ and $\operatorname{Var}(X_i)>0$ for $i=1,2$. Using b) we conclude that this is equivalent to $\min\{\lambda_1,\lambda_2\}>2$ as in this case the second condition is automatically satisfied.
d) Since $\min\{\lambda_1,\lambda_2\}=3$, exercise c) shows that the linear correlation is well-defined. Hence, Hoeffding's identity implies $\rho\in[\rho_{\min},\rho_{\max}]$ whereas the the minimal, maximal linear correlation is attained if Y_1 and Y_2 are coupled by the counter-monotonicity, comonotonicity copula $W(u,v)=(u+v-1)_+, M(u,v)=\min\{u,v\}$, respectively (1 Pts). Hence, we have $\rho_{\min}=\rho(Y_1,Y_2)$ and $\rho_{\max}=\rho(Y_1,Y_2)$ in the respective cases. To calculate $\rho_{\min},\rho_{\max}$ explicitly, we need $\mathbb{E}[Y_1],\mathbb{E}[Y_2],\operatorname{Var}(Y_1),\operatorname{Var}(Y_2)$ and $\operatorname{Cov}(Y_1,Y_2)$.
Using the calculations from b), one easily obtains $\mathbb{E}[Y_1]=\frac{3}{2},\mathbb{E}[Y_2]=\frac{4}{3},\operatorname{Var}(Y_1)=\frac{3}{4},\operatorname{Var}(Y_2)=\frac{2}{9}$
Finally, to compute $\operatorname{Cov}(Y_1,Y_2)$ we need $\mathbb{E}[Y_1Y_2]$. If Y_1,Y_2 are coupled by the countermonotonicity, comonotonicity copula, respectively, we know $(Y_1,Y_2)\overset{(d)}{=}\left(q_{Y_1}(U),q_{Y_2}(1-U)\right)$, $(Y_1,Y_2)\overset{(d)}{=}\left(q_{Y_1}(U),q_{Y_2}(U)\right)$ for some $U\sim\operatorname{Unif}(0,1)$ and q_{Y_1},q_{Y_2} are quantile functions of Y_1,Y_2. By inverting the distrubtion functions of Y_1,Y_2 we obtain $q_{Y_1}(u)=(1-u)^{-1/3}$ $q_{Y_2}(v)=(1-v)^{-1/4}$ for all $u,v\in(0,1)$.
Hence, we obtain $\mathbb{E}[Y_1Y_2]=\mathbb{E}[h(Y_1,Y_2)]=\mathbb{E}\left[h\left(q_{Y_1}(U),q_{Y_2}(1-U)\right)\right]=\int_0^1(1-x)^{-1/3}x^{-1/4}dx$ and $\mathbb{E}[Y_1Y_2]=\mathbb{E}[h(Y_1,Y_2)]=\mathbb{E}\left[h\left(q_{Y_1}(U),q_{Y_2}(U)\right)\right]=\int_0^1(1-x)^{-(1/3+1/4)}dx=\frac{12}{5}$ where $h:\mathbb{R}^2\to\mathbb{R}$ is given by $h(x,y)=xy$ for all $x,y\in\mathbb{R}$. Recalling that the Beta function is given by $B(x,y)=\int_0^1 t^{x-1}(1-t)^{y-1}dy$. So $\mathbb{E}[Y_1Y_2]=B(3/4,2/3)$
Moreover, we have $\operatorname{Cov}(Y_1,Y_2)=\mathbb{E}[Y_1Y_2]-\mathbb{E}(Y_1)\mathbb{E}(Y_2)=\mathbb{E}[Y_1Y_2]-2$
Using $\rho(Y_1,Y_2)=\operatorname{Cov}(Y_1,Y_2)/\sqrt{\operatorname{Var}(Y_1)\operatorname{Var}(Y_2)}$ we finally obtain $\rho_{\min}=\rho(Y_1,Y_2)=\frac{B(3/4,2/3)-2}{\sqrt{6/36}}$ and $\rho_{\max}=\rho(Y_1,Y_2)=\frac{2\sqrt{6}}{5}$

Ex4 Quantile Transformation

Construct a two-dimensional random vector (X_1,X_2) such that (i) $X_i\sim\operatorname{Exp}(\lambda_i)$ for $\lambda_i>0, i=1,2$, and (ii) $\operatorname{VaR}_\alpha(X_1+X_2)=\operatorname{VaR}_\alpha(X_1)+\operatorname{VaR}_\alpha(X_2)$ for all $\alpha\in(0,1)$

Solution: Let $U\sim\operatorname{Unif}(0,1)$ and let $q_i(u)=-\frac{1}{\lambda_i}\log(1-u), i=1,2$. Since, q_i is the quantile function of $\operatorname{Exp}(\lambda_i)$ distribution, we have by the quantile transformation lemma that $q_i(U)\sim\operatorname{Exp}(\lambda_i)$. Setting $X_i=q_i(U)$, we have $X_1+X_2=q_1(U)+q_2(U)=(q_1+q_2)(U)$. The quantile transformation then implies that q_1+q_2 is the quantile function of X_1+X_2. We thus have by the definition of VaR that $\operatorname{VaR}_\alpha(X_1+X_2)=(q_1+q_2)(\alpha)=q_1(\alpha)+q_2(\alpha)=\operatorname{VaR}_\alpha(X_1)+\operatorname{VaR}_\alpha(X_2)$

Ex5. Exchangeability

Let X be a d-dimensional random vector with a $t_d(\nu,0,\Sigma)$-distribution for $d\geq 2,\nu>0$ and a positive definite $d\times d$-matrix Σ. Are the components of X exchangeable?

Solution: We say that a random vector $X=(X_1,\ldots,X_d)$ is exchangeable if $(X_1,\ldots,X_d)\overset{(d)}{=}\left(X_{\pi(1)},\ldots,X_{\pi(d)}\right)$ for any permutation π of $\{1,\ldots,d\}$.

Since every permutation of $\{1,\ldots,d\}$ can be represented by a $d\times d$-matrix P with $P_{ij}=\mathbb{1}_{\{\pi(i)=j\}}$ for all $i,j\in\{1,\ldots,d\}$ and since we know from the lecture that t distribution is (as a normal variance mixture) closed under affine transformations, we have that $Y:=PX\sim t_d\left(\nu,0,P\Sigma P^\top\right)$. It is therefore enough to find a positive definite $d\times d$-matrix Σ, such that $P\Sigma P^\top\neq\Sigma$. Without loss of generality, take Σ diagonal with $\Sigma_{11}>\Sigma_{22}$ and P a permutation matrix corresponding to a permutation that swaps the first and second component of X (that is a matrix obtained by swapping the first and the second row of the $d\times d$ identity matrix). In that case it follows that $\left(P\Sigma P^\top\right)_{11}=\Sigma_{22}$ and $\left(P\Sigma P^\top\right)_{22}=\Sigma_{11}$, that is, $P\Sigma P^\top\neq\Sigma$.

Ex6. Quantile Transportation to Prove AVaR Eq

Let X be a random variable such that $\mathbb{E}[|X|]<\infty$. Show that $\operatorname{AVaR}_\alpha(X)=\operatorname{VaR}_\alpha(X)+\frac{1}{1-\alpha}\mathbb{E}[(X-\operatorname{VaR}_\alpha(X))_+]$ for all $\alpha\in(0,1)$.

Solution:

$$\begin{aligned}
&\operatorname{AVaR}_\alpha(X)\\
&=\frac{1}{1-\alpha}\int_\alpha^1\operatorname{VaR}_u(X)du\\
&=\operatorname{VaR}_\alpha(X)+\frac{1}{1-\alpha}\int_\alpha^1(\operatorname{VaR}_u(X)-\operatorname{VaR}_\alpha(X))du\\
&=\operatorname{VaR}_\alpha(X)+\frac{1}{1-\alpha}\int_0^1(\operatorname{VaR}_u(X)-\operatorname{VaR}_\alpha(X))_+\mathbb{1}_{(\alpha,1)}(\\
&=\operatorname{VaR}_\alpha(X)+\frac{1}{1-\alpha}\mathbb{E}_U\left[\left(q_U^-(X)-\operatorname{VaR}_\alpha(X)\right)_+\right]
\end{aligned}$$

where $U\sim\operatorname{Unif}(0,1),\alpha\in(0,1)$. Using the quantile transformation theorem, we obtain $\operatorname{AVaR}_\alpha(X)=\operatorname{VaR}_\alpha(X)+\frac{1}{1-\alpha}\mathbb{E}[(X-\operatorname{VaR}_\alpha(X))_+]$ for all $\alpha\in(0,1)$